AF $ 2200

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Pastan et al.
Application No. 09/763,393
Filed: July 30, 2001
Confirmation No. 5265
For: PAGE-4, AN X-LINKED GAGE-LIKE GENE EXPRESSED IN NORMAL AND NEOPLASTIC PROSTATE, TESTIS AND UTERUS, AND USES THEREFOR
Examiner: Minh-Tam Davis
Art Unit: 1642
Attorney Reference No. 4239-61541-01

CERTIFICATE OF MAILING

I hereby certify that this paper and the documents referred to as being attached or enclosed herewith are being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: MAIL STOP APPEAL BRIEF – PATENTS, COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450 on the date shown below.

Attorney or Agent for Applicant(s) _____

Date Mailed December 8, 2006

MAIL STOP APPEAL BRIEF - PATENTS
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450

TRANSMITTAL LETTER

Enclosed for filing in the application referenced above are the following:

- ☒ Appeal Brief and 11 references listed on Evidence Appendix
- ☒ Fee for Filing Appeal Brief - $500.00
- ☒ Fee for Two-month Extension of Time - $450.00
- ☒ A check in the amount of $950.00 to cover the above-listed fees

- ☒ Applicants petition for an extension of time for 2 month(s). If an additional extension of time is required please consider this a petition therefor.
- ☒ The Director is hereby authorized to charge any additional fees that may be required, or credit over-payment, to Deposit Account No. 02-4550. A copy of this sheet is enclosed.
- ☒ Please return the enclosed postcard to confirm that the items listed above have been received.

Respectfully submitted,

KLARQUIST SPARKMAN, LLP

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 595-5300
Facsimile: (503) 595-5301

By _____
Susan Alpert Siegel, Ph.D.
Registration No. 43,121

cc: Docketing

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Pastan et al.
Application No. 09/763,393
Filed: July 30, 2001
Confirmation No. 5265
For: PAGE-4, AN X-LINKED GAGE-LIKE GENE EXPRESSED IN NORMAL AND NEOPLASTIC PROSTATE, TESTIS AND UTERUS, AND USES THEREFOR
Examiner: Minh-Tam Davis
Art Unit: 1642
Attorney Reference No. 4239-61541-01

CERTIFICATE OF MAILING

I hereby certify that this paper and the documents referred to as being attached or enclosed herewith are being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: MAIL STOP APPEAL BRIEF – PATENTS COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450 on the date shown below.

Attorney or Agent for Applicant(s)

Date Mailed December 8, 2006

MAIL STOP APPEAL BRIEF - PATENTS
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450

APPEAL BRIEF

This is an appeal brief filed under 37 C.F.R. §41.37. Applicants believe the Notice of Appeal was received by the U.S. Patent and Trademark Office (USPTO) on August 9, 2006, making the Appeal Brief due on October 9, 2006. A two-month extension for the period for response is hereby requested, making the Appeal Brief due on December 9, 2006.

Real Party in Interest

The real party in interest is the United States of America, represented by the Secretary, Department of Health and Human Services, the assignee of the present application.

Related Appeals and Interferences

There are no related Appeals or Interferences.

12/11/2006 EAREGAY1 00000081 09763393
01 FC:1402 500.00 OP

12/11/2006 EAREGAY1 00000081 09763393
02 FC:1252 450.00 OP

Status of Claims

Claims 1, 2, 4, 6-8, 14-15, 17-18 and 53-57 are pending. Claims 9-12, 16, 19-52 and 58-60 are withdrawn from consideration. Claims 3, 5, 13 and 19-52 have been canceled. Claims 1, 2, 4, 6-8, 14-15, 17-18, and 53-57 have been rejected, and are appealed.

Status of Amendments

No amendments have been filed subsequent to the final rejection.

Summary of Claimed Subject Matter

The claims on appeal relate to isolated polypeptides comprising (1) an amino acid sequence set forth as SEQ ID NO: 1 (claims 1 and 53) (see, for example, Fig. 1 of the specification, page 8, lines 17-37, and SEQ ID NO: 1 of the sequence listing) or (2) polypeptides comprising (claim 1) or consisting of (claim 54) 8 to 11 contiguous amino acids of SEQ ID NO: 1, wherein the polypeptide binds major histocompatibility complex (MHC) I (see, for example, page 7, line 32 to page 8, line 25, and page 20, line 1 to page 23, line 11). In one embodiment, the polypeptides bind HLA-A1, HLA-A2.1, HLA-A3.2, HLA-A4.1 or HLA-A11.2 (claim 56) (see, for example, page 20, lines 10-19 and on page 21, lines 13-29). In another embodiment, the polypeptides are conjugated to a lipid (claim 57, see for example page 23, lines 18-21). In a further embodiment (claims 6-8) immunogenic compositions including these polypeptides are disclosed (see page 20, line 1 to page 23, line 32 and page 33, line 10 to page 36, line 15). These compositions can include a stabilizing detergent, a micelle-forming agent, and an oil (claims 7-8) (see, for example, page 34, line 17 to page 35, line 18). In another embodiment, methods are provided for inhibiting the growth of a malignant cell expressing SEQ ID NO: 1 by (a) culturing cytotoxic T lymphocytes (CTLs) or CTL precursor cells with the claimed polypeptides to activate the CTLs or CTL precursors and (b) contacting the malignant cell with the activated CTLs or CTLs matured from the CTL precursors (claim 14) (see, for example, page 24, line 1 to page 25, line 10).

Grounds of Rejection to be Reviewed on Appeal

(1) Rejection of claims 1-2, 4, 6-8, 15, 17-18 and 54-57 under 35 U.S.C. § 112, first paragraph, as allegedly there is insufficient written description for polypeptides of eight to ten amino acids of SEQ ID NO: 1 that bind MHC.

(2) Rejection of claims 1-2, 4, 6-8, 17-18 and 53-57 under 35 U.S.C. § 101 as allegedly there is no utility for the claimed polypeptides.

(3) Rejection of claims 1-8, 14-15, 17-18 and 53-57 under 35 U.S.C. § 112, first paragraph, as allegedly not being enabled by the specification.

Argument

(1) There is sufficient written description for a peptide comprising eight to ten consecutive amino acids of SEQ ID NO: 1 (claims 1-8, 14-15) or consisting of (claims 54-57) eight to ten consecutive amino acids of SEQ ID NO: 1 that bind MHC.

(A) Claims 2, 7 and 17

The final Office action states that there is insufficient written description for polypeptides comprising eight to ten amino acids of SEQ ID NO: 1. All of the rejections in the final Office action are directed to polypeptides including eight to ten amino acids of SEQ ID NO: 1. No arguments were presented in the final Office action that indicate that the U.S. PTO believes that there is insufficient written description for polypeptides comprising SEQ ID NO: 1 (a full-length PAGE4 amino acid sequence). Indeed, claim 53, directed to a polypeptide comprising an amino acid sequence set forth as SEQ ID NO: 1 is not included in this rejection. *Thus, the U.S. Patent and Trademark Office has acknowledged that there is sufficient written description for polypeptides comprising SEQ ID NO: 1.*

Claims 2, 7 and 17 depend from claim 53. Claims 2 and 7 are directed to immunogenic compositions including the polypeptide of claim 53; Claim 17 is directed to methods for inhibiting the growth of a malignant cell using the immunogenic composition of claim 2. The final Office action does not provide any indication that there is insufficient written description

for either (a) pharmaceutical/immunogenic compositions or (b) methods of inhibiting the growth of a malignant cell. Thus, it is the Applicants understanding that there is sufficient written description for the claimed compositions and methods of use.

Since the final Office action did not provide any basis for rejecting claims 2, 7, 17 and 53 as lacking written description, Applicants do not have specific rejections that can be addressed in this Appeal Brief. The rejection of claims 2, 7 and 17 under 35 U.S.C. § 112, first paragraph should be reversed because the examiner has not established a basis for the rejection.

(B) Claims 1, 4, 6, 8, 18, and 54-57

Claims 1, 4, 6, 8, 18 and 54-57 are directed to polypeptides comprising (claim 1) or consisting of (claim 54) an amino acid sequence of 8 to 10 consecutive amino acids of SEQ ID NO: 1 and their use. All of the claims are limited to polypeptides (or the use of polypeptides) that have a specified function, namely that the polypeptide binds major histocompatibility complex (MHC) class I.

The final Office action asserts that (i) the open language "comprising" results in claiming too many unknown sequences that include 8 to 11 amino acids of SEQ ID NO: 1; and (ii) that the claimed genus of peptides cannot share a common structure, because "the different regions of SEQ ID NO: 1 are structurally different" (see page 3 of the final Office action). The final Office action also asserts that binding to MHC class I is not a critical function (see page 4 of the final Office action). The final Office action also alleges that the structure of the peptides of 8-11 amino acids in length that could induce cytotoxic T lymphocytes (also known as "T cells" or "CTLs") that could lyse primary cancer cells is not disclosed (see page 5 of the final Office action). The final Office action alleges that as there is no common structure and no critical function and there is no correlation between function and sequence there must be insufficient written description for the claimed polypeptides (see page 4 of the final Office action). The final Office action further alleges that there is not sufficient description of the subgenus of peptides of 9 to 10 amino acids in length that include specified anchor residues and bind MHC. The Office action cites to Visseren et al. and Grey et al. in support of these assertions.

The final Office action acknowledges receipt of the Declaration of Dr. Pastan under 37 C.F.R. § 1.132 (hereinafter "the Declaration"), which was submitted in response to the first

Office action (copy enclosed). The Office action acknowledges that the Declaration documents that the claimed polypeptides could be made and used to lyse tumor cells. However, the final Office action asserts that (iii) as the Declaration is post-filing, it cannot be used to obviate a rejection under 35 U.S.C. § 112, first paragraph for a lack of written description. The Office action asserts that since the Declaration does not provide description of a class of peptides that share a common structure, it does not obviate the rejection (see page 8 of the final Office action). Points (i)-(iii) are addressed individually below.

(i) There is sufficient support for polypeptides "comprising" 8 to 11 amino acids of SEQ ID NO: 1.

Only claim 1 is directed to polypeptides comprising 8 to 11 amino acids of SEQ ID NO: 1 that bind MHC. Claims 54, and dependent claims therefrom, are directed to polypeptides "consisting of" 8 to 11 contiguous amino acids of SEQ ID NO: 1.[1] There is considerable guidance for polypeptides comprising 8 to 11 amino acids of SEQ ID NO: 1, which is 102 amino acid in length. The full length sequence itself provides multiple instances of subsequences that would comprise the 8 to 11 amino acid sequence. The conjugation of polypeptides to other agents, (for example, a lipid) is described in the specification, such as on page 23, lines 21-21. Pharmaceutical compositions including these polypeptides are described in the specification, for example on page 33, line 10 to page 35, line 19.

An objective standard is provided in MPEP §2163.02 for determining compliance with the written description requirement. Specifically, the description must "clearly allow persons of ordinary skill in the art to recognize that he or she invented what is claimed." *In re Gosteli*, 872 F.2d 1008, 1012, 10 USPQ2d 1614, 1618 (Fed. Cir. 1989). Under *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 1563-64, 19 USPQ2d 1111, 1117 (Fed. Cir. 1991), the test for sufficiency of support in a patent application is whether the disclosure of the application relied upon "reasonably conveys to the artisan that the inventor had possession at that time of the later claimed subject matter." *Ralston Purina Co. v. Far-Mar-Co., Inc.*, 772 F.2d 1570, 1575, 227

[1] Claim 53 is directed to polypeptides comprising SEQ ID NO: 1 (full-length protein), and is not included in the rejection as set forth in the final Office action. Thus, Applicants believe that the Examiner has raised this rejection only with regard to claim 1.

USPQ 177, 179 (Fed. Cir. 1985) (quoting *In re Kaslow*, 707 F.2d 1366, 1375, 217 USPQ 1089, 1096 (Fed. Cir. 1983)). In the present application, there clearly is sufficient written description for the claimed polypeptides.

In order for the specification to provide sufficient written description, one of skill in the art must be able to produce a polypeptide including 8 to 11 amino acids of SEQ ID NO: 1 that bind MHC using the guidance provided by the specification. MPEP § 2163 (3)(a) states:

> What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail. See *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d at 1384, 231 USPQ at 94. If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met. See, e.g., *Vas-Cath*, 935 F.2d at 1563, 19 USPQ2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 USPQ 391, 395 (CCPA 1972).

It was conventional and routine for one of skill in the art to produce polypeptides including a specified amino acid sequence at the time the present application was filed. A polypeptide that includes 8 to 11 consecutive amino acids of PAGE4 (SEQ ID NO: 1) can be chemically synthesized by standard methods; automated machinery for chemical synthesis was available at the time the present application was filed. In addition, polypeptides could also be produced using molecular genetic techniques, such as by inserting a nucleic acid encoding 8 to 11 consecutive amino acids of SEQ ID NO: 1 into an expression vector, introducing the expression vector into a host cell, and isolating the polypeptide. Molecular techniques to produce and isolate proteins are well known in the art. These techniques are described in standard textbooks, for example, *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and are disclosed in the specification (see, for example, page 27, line 31 to page 31, line 1). Thus, the specification clearly provides sufficient written description of a molecular biologist to produce polypeptides including 8 to 11 amino acids of SEQ ID NO: 1.

(ii) The specification provides a defined structure and a critical function for the claimed polypeptides

Structural information is provided in the specification for polypeptides having an amino acid sequence set forth as SEQ ID NO: 1. The complete amino acid sequence of PAGE4 is

shown in SEQ ID NO: 1 and FIG. 5 of the specification. Functional domains of SEQ ID NO: 1 are disclosed in the specification on pages 14-15, which describes similarity of the amino acid sequence set forth as SEQ ID NO: 1 (PAGE4) with GAGE and MAGE proteins. The specification describes the presence of RGD motifs (which are disclosed to be involved in protein-protein interactions) in SEQ ID NO: 1. Thus, the structure of SEQ ID NO: 1 is clearly set forth in the specification.

In addition, the specification clearly describes immunogenic epitopes of PAGE4 in sufficient detail to clearly convey to the public what was invented and to put the public in possession of what the applicant claims as the invention. For example, immunogenic peptides, such as peptides that bind MHC are disclosed in the specification on page 7, line 35 to page 8, line 25, and on page 20, line 1 to page 22, line 5. The specification also discloses that epitopes of use are 8-10 amino acids in length and have anchoring residues, such as at positions 2 and 9 of the PAGE4 polypeptide (see page 20, lines 33-34). Specific types of PAGE4 polypeptides that are of specific use are disclosed. For example, it is disclosed that the PAGE4 polypeptides can be 9 or 10 amino acids in length and can include binding motifs for HLA-A2 (see, for example, page 8, lines 30-37; page 20, to page 21, line 2, and page 21, lines 15-19).

The final Office action alleges that since binding to Major Histocompatibility Complex (MHC) class I is not a critical feature of the claimed peptides (see page 6 of the final Office action), and because there is no correlation between structure and function disclosed in the specification, there must be insufficient written description for the claimed peptides. Applicants respectfully disagree.

In order to produce a T cell response, a peptide must be presented by MHC. This is depicted in Figure 4-1 of Abbas et al., (from "Cellular and Molecular Immunology," 4th edition, W. B. Saunders Co. Philadelphia, 2000 (1st edition, 1991), page 64, copy attached); this figure illustrates the binding of peptides to MHC. The importance of anchor residues peptide binding to MHC class I is shown in Fig. 4-1 of Abbas et al. In addition, Abbas et al. confirms that a critical function required for peptides that induce a T cell response is binding to MHC class I molecules. For example, Abbas et al. states (page 71) "....we now understand the physicochemical characteristics of peptide-MHC interactions in considerable detail." Abbas et al. also states (page 72) "[t]he peptides that bind to MHC molecules share structural features that promote this interaction." Contrary to the examiner's assertion, one of skill in the art clearly

understands that binding to MHC is a critical feature to promote a T cell response. In addition, one of skill in the art can readily identify defining characteristics of a peptide that binds MHC, namely the anchor residues.

MPEP § 2163 states:

> For some biomolecules, examples of identifying characteristics include a sequence, structure, binding affinity, binding specificity, molecular weight, and length. Although structural formulas provide a convenient method of demonstrating possession of specific molecules, other identifying characteristics or combinations of characteristics may demonstrate the requisite possession. For example, [>]disclosure of an antigen fully characterized by its structure, formula, chemical name, physical properties, or deposit in a public depository provides an adequate written description of an antibody claimed by its binding affinity to that antigen. *Noelle v. Lederman*, 355 F.3d 1343, 1349, 69 USPQ2d 1508, 1514 (Fed. Cir. 2004).

Thus, to provide sufficient written description for a biomolecule, the specification should provide a sequence, structure, binding affinity, binding specificity, molecular weight, and length. In the present application, the Applicants have provided the sequence, namely the complete amino acid sequence set forth as SEQ ID NO: 1. A polypeptide with this sequence has a defined molecular weight.[2] Moreover, the claimed peptides are of a specific length, namely of 8 to 11 amino acids in length, or 9 to 10 amino acids in length. In addition, the claimed polypeptides have a defined binding specificity and affinity: they specifically bind MHC class I.

The specification provides additional written description for the claimed polypeptides. The specification discloses that peptides of use have specific anchoring residues in the second position (A, L, I, V, M or S) and a positively charged amino acid at the position nine (see page 20, line 20 to page 21, line 2). The selection of binding motifs that bind HLA-A2 (the specified function of binding MHC class I) is disclosed on page 28, line 25 to page 29, line 29. In addition, biological methods to test whether a specific epitope is immunogenic are provided (for example, see page 8, lines 1-4 and page 21, lines 3-12 and lines 20-29). Specific types of PAGE4 polypeptides that are of specific use are disclosed. For example, the specification discloses that the PAGE4 polypeptides can be 9 or 10 amino acids in length and can include binding motifs for HLA-A2 (see, for example, page 8, lines 30-37; page 20, to page 21, line 2, and page 21, lines 15-19).

[2] Each amino acid has a known molecular weight. To determine the molecular weight of a polypeptide of 102 amino acids, such as SEQ ID NO:1, the individual weights of each of the 102 amino acids is added together.

Computer programs for predicting MHC binding were well known to those of skill in the art at the time the provisional application was filed (see for example, Parker et al., J Immunol 152:163-75, 1994; Rammensee et al., Immunogenetics 50: 213-219, 1999, copies enclosed). These programs are available (without charge) on the internet. Considerable supporting information is also provided in the specification, for example methods and computer based programs for predicting MHC binding motifs (immunogenic epitopes) are disclosed in the specification (for example, see page 8, lines 12-25 and page 21, line 34 to page 22, line 5). Thus, given the guidance provided by the specification, one of skill in the art could readily produce polypeptides of a defined structure (8 to 11 consecutive amino acids of SEQ ID NO: 1 or 9 to 10 consecutive amino acids of SEQ ID NO: 1) that have the claimed function (they bind MHC class 1).

Grey et al. and Visseren et al. support the conclusion that adequate guidance is provided in the specification

The Office action cites Grey et al. (PCT Publication No. WO 94/010127 A1, copy enclosed) as teaching that charged amino acids in non-anchor positions can affect MHC binding, and thus allegedly documenting the insufficiency of the written description of the present specification. Applicants respectfully disagree, and suggest that Grey et al. instead documents that one of skill in the art, given the amino acid sequence of a full-length polypeptide and an algorithm (such as a computer program), can readily produce peptides that bind MHC and induce an immune response.

Grey et al. discloses that peptides that bind MHC are nine to ten amino acids in length and include specified anchor residues at positions 9 and 10 (see page 3, first paragraph). These peptides bind MHC, specifically HLA 2.1. Exemplary peptides are provided from a variety of antigens, including prostate specific antigen and viral antigens, amongst others. Grey et al. describe that superior binding is achieved with a leucine at position 2 and a valine at position 9 (see page 40), and describes the importance of the anchor residues. Grey et al. describes a "motif library approach" to determine the importance of the non-anchor residues (see page 44), and discloses that predictable trends can be assessed from the motif library approach (see pages 45-59). Grey discloses that an algorithm (such as a computer program) can be designed to predict MHC binding (see pages 65-68). The reference further states "the results of Examples 6 and 7

[the algorithm] indicate that an algorithm can be used to select peptides that bind to HLA-A2.1 sufficiently to have a high probability of being immunogenic." Grey et al. describes the use of an algorithm to predict specific papilloma virus peptides that will bind MHC, and then provide data documenting that four human papilloma virus peptides that bound MHC were generated (page 95, lines 1-20). All of these peptides were shown to be immunogenic (see page 97, lines 19-30). Thus, Grey et al. provide confirmation that the disclosure of an amino acid sequence of a full-length polypeptide and an algorithm (such as a computer prediction program) is sufficient written description for one of skill in the art to produce peptides that bind MHC.

The Office action also refers to Visseren et al., Int. J. Cancer 73: 125-30, 1997 (copy enclosed), as showing that a full-length polypeptide sequence and specified physical characteristics (such as anchor residues) cannot provide sufficient written description for peptides that bind MHC. Applicants respectfully disagree.

Visseren et al. discloses that the binding of a peptide to MHC class I is determined by "its length and the fit of the amino acid side chains of the peptides into pockets of the MHC molecule" (page 126, second column, under "Results"). Visseren et al. teach that using the sequence of MAGE-2, and using a computer program, 22 peptides were generated that fit the binding motif. Of these peptides 14 competed for the binding of a reference peptide, and only 7 peptides were incapable of binding a reference peptide. Additional experiments performed by Visseren et al. document that the MAGE-2 peptides predicted to bind MHC by the computer program were immunogenic in transgenic mice. Thus, the data presented in Visseren et al. document that peptides predicted to bind MHC generally were immunogenic, and suggest that epitope based vaccines including some of the predicted peptide sequences can be used for prevention and treatment of tumors (see Visseren et al., page 129). Thus, it is clear from Visseren et al. that one of skill in the art, given an amino acid sequence, guidance on anchor residues, and guidance on computer prediction programs of use to predict peptides that bind MHC, could readily identify, produce and use peptides that bind MHC. Applicants' specification provides all of this information and therefore satisfies the written description requirement.

The Office action cites to sections of Visseren et al. that show that, for MAGE-2, of the 22 computer-predicted peptide sequences there were "several" that bound at 4 ° C that did not form highly stable complexes at 37 ° C. Visseren et al. provides this information as a test to

indicate which peptides will be effective *in vivo*. However, Visseren et al. states that all the peptides that did form stable complexes with MHC at 37 ° C produced an immune response in mice (see page 127, first paragraph).

Applicants do not deny that some of the MAGE-2 peptides produced by Visseren et al. did not bind MHC with the high affinity. However, Visseren et al. document that techniques are available that are routine to one of skill in the art to (1) produce peptides predicted to bind MHC, (2) perform confirmatory test to demonstrate MHC binding, and to (3) confirm that these peptides are immunogenic.

There is no requirement under 35 U.S.C. § 112 that all of the claimed peptides must possess the same degree of utility (*In re Gardener* 475 F2d 1389, 177 USPQ 396 (CCPA 1973)). The Federal Circuit has confirmed that that the written description "requirement may be satisfied if, in the knowledge of the art, the disclosed function is sufficiently correlated to a particular, known structure" (*Amgen Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313, 1319 (Fed. Cir. 2003)). Indeed, Grey et al. and Visseren et al. provide evidence that, the specification provides adequate guidance, as it provides the amino acid sequence of the full-length PAGE4 polypeptide (SEQ ID NO: 1), information on the characteristics that regulate binding to MHC, and information on computer programs to make and use the claimed polypeptides.

(iii) The Declaration of Dr. Pastan documents the sufficiency of the specification

The specification clearly describes the claimed PAGE4 peptides and the use of these PAGE4 polypeptides to induce a cytotoxic T cell response. The description provided in the specification for PAGE4 polypeptides of 8 to 11 amino acids in length of SEQ ID NO: 1 that bind MHC class I is discussed above.

The sufficiency of the specification was further supported by the Declaration of Dr. Pastan, which was submitted to the U.S. PTO on March 15, 2006. In this declaration, Dr. Pastan discloses that, using the information disclosed in the specification, the primary amino acid sequence of human PAGE4 was analyzed for consensus motifs for novel HLA-A2 binding peptides using a computer program. Specifically, the amino acid sequence of PAGE4 was scanned for matches to consensus motifs for HLA-A2 binding peptides using the algorithm from the BioInformatics and Molecule Analysis Section of NIH (BIMAS) (Parker et al., J Immunol 152:163-75, 1994, copy enclosed), which ranks potential MHC binding peptides according to the

predictive one-half-time dissociation of peptide/MHC complexes. The HLA-A2 allele was chosen because it is the most commonly expressed class I allele. Nine-mer and ten-mer peptides were synthesized if they conformed to the respective consensus motif (binding of immunogenic peptides to this allele is disclosed in the specification, see for example page 20, lines 10-19). Three native PAGE4 peptides of nine to ten consecutive amino acids of PAGE4 (SEQ ID NO: 1) were produced that were predicted to have high binding affinity. Binding of PAGE4 peptides to HLA-A2 molecules was experimentally evaluated by the up-regulation of HLA-A2 expression on T2 cells as demonstrated by flow cytometry (Njiman et al., Eur J Immunol 23:1215-9, 1993). Two of three native peptides (P16, P84) bound to HLA-A2 molecules in the this assay. Studies confirmed that the peptide-MHC complex was stable for the two peptides with high affinity (P16 and P84). Thus, three polypeptides consisting of 9 to 10 amino acids in length of SEQ ID NO: 1 were produced that bound MHC class 1 using the guidance provided in the specification; two of these polypeptides were experimentally confirmed to stably bind MHC class I with a high affinity.

The claimed peptides bind MHC class I. Although not specifically recited in the pending claims, the use of peptides that bind MHC class I is to activate cytotoxic T lymphocytes ("CTLs"). The specification provides additional description of methods that use PAGE4 polypeptides that bind MHC class I to activate CTL *ex vivo* (see page 24, line 1 to page 25, line 10). The specification also discloses methods for the induction of a CTL response in vivo, using the claimed polypeptides (see page 33, line 10 to page 35, line 19).

The Declaration of Dr. Pastan also describes experiments that were performed to confirm that T-cell lines could be generated from peripheral blood mononuclear cells (PBMC) from prostate cancer patients using polypeptides consisting of 9 to 10 amino acids of SEQ ID NO: 1 that bound MHC class I. PAGE4-specific T-cell lines were generated from a prostate cancer patient (termed patient "A") using two polypeptides (P16 and P84). High levels of cytokine production (IFN-γ, used to measure T cell activation) were observed when the T-cell lines were stimulated with both the P16 and P84 peptides, although a higher level of IFN-γ production was observed for the T-A-P16 T-cell line (produced with P16) for the T-A-P84 T-cell line (produced with P84). These results confirmed that cytotoxic T cells could be activated using the claimed polypeptides. A single PAGE4 peptide (P16) was then selected for additional confirmatory

studies. The results demonstrated that CTLs generated using P16 could lyse prostate tumor cells that express PAGE4 (SEQ ID NO: 1).

These studies described in the Declaration demonstrate that immunogenic fragments of PAGE4 can be produced and used to activate lymphocytes, as described in the specification. In addition, the results demonstrate that immunogenic PAGE4 fragments can be used to induce an immune response that results in the lysis of prostate cancer cells. The results presented (see points 2-7 of the Declaration) documented that *using the description provided by the specification,* Dr. Pastan and his colleagues (who are of skill in the art) were able to generate immunogenic PAGE4 polypeptides, and use these polypeptides to activate cytotoxic T cells.

However, the Office action states that that as the data in the declaration presents post-filing date evidence, the data in the Declaration could not be used for obviating the rejection. Applicants respectfully disagree with this refusal to consider the Declaration. In order to determine the sufficiency of the guidance provided by the present specification to produce the claimed peptides, one of skill in the art must use the description included in the patent application as filed to determine if it is sufficient to direct one of skill in the art to produce the claimed peptides. Thus, any evidence documenting the sufficiency of the description to direct one of skill in the art to make (or use) the claimed peptides must, by definition, be obtained after the application is filed. Thus, it is inappropriate for the U.S. PTO to disregard the Declaration which demonstrates that the specification satisfies 35 U.S.C. § 112, first paragraph.

Conclusion

The specification clearly provides sufficient written description for the pending claims. The test for sufficiency, namely that the specification "reasonably conveys to the artisan that the inventor had possession at that time of the claimed subject matter" is supported by the Declaration of Dr. Pastan. In addition, the prior art cited in the final Office action (Grey et al. and Visseren et al.) support the conclusion that one of skill in the art, given the amino acid sequence of a full length polypeptide and guidance on functional characteristics for binding MHC can readily produce peptides of 8-11 amino acids in length that bind MHC and can be used to produce an immune response. Thus, the rejection of claims -2, 4, 6-8, 15, 17-18 and 54-57 under 35 U.S.C. § 112, first paragraph should be withdrawn.

(2) There is a specific, substantial and credible utility for the compositions of claims 1-8, 14-15, 17-18 and 53-57.

The final Office action alleges that there is no utility for the claimed polypeptides. Specifically, the final Office action alleges that although the specification discloses that the claimed polypeptides can be used to detect prostate cancer in subjects and discloses that the claimed polypeptides can be used to produce an immune response to cancers expressing PAGE4 (SEQ ID NO: 1), there is no *proven* utility for the claimed peptides. The final Office action acknowledges that there is a description of this use and confirms that the specification documents the expression of mRNA encoding PAGE4 in prostate cancer. However, the final Office action asserts that, as there is no correlation between mRNA expression and protein expression for some polypeptides, that data on the detection of the presence of mRNA encoding PAGE4 in tumors is insufficient to support either disclosed use. The final Office action acknowledges that both the Declaration of Dr. Pastan and Iavraone et al. (copy enclosed) document the utility of the claimed polypeptides. However, the final Office action discounts this data as being post-filing. The final Office action further asserts that as cancer treatment is unpredictable there is no utility for the claimed proteins. In addition, the final Office action states that the specification gives only an "invitation to experiment" such that there is "no specific asserted utility" (see page 19 of the Office action). Applicants respectfully disagree.

To satisfy the requirements of 35 U.S.C. § 101, an invention must be useful, and the specification must include specific information to make it apparent to one of skill in the art that the use is "specific," "substantial" and "credible." The present application discloses a polypeptide including or consisting of the amino acid sequence set forth as SEQ ID NO: 1, or a polypeptide including or consisting of 8 to 11 amino acids of SEQ ID NO: 1. The specification discloses that these polypeptides are of use to (1) detect and (2) treat cancer, such as prostate cancer. This use is "specific" for claimed polypeptides, is "substantial" as is the "real world" use of cancer treatment, and is "credible" as the production of an immune response against tumor antigens is known to be of use for cancer treatment (as evidenced by Visseren et al., discussed above). These uses are clear and defined, contrary to the assertion in the final Office action that they are an "invitation to experiment" (which is not a proper test for utility).

1. Discordance between mRNA level and protein levels for unrelated genes does not negate the disclosed utility for the claimed polypeptides

The final Office action asserts that "protein levels cannot be predictably correlated with steady-state mRNA levels." The Office action alleges that, since the presence of an mRNA does not always correlate with the presence of a protein, a disclosed use for a protein (either to detect prostate cancer or to treat prostate cancer) cannot be supported only by data showing mRNA expression. In support of this assertion the Office action cites:

1. Brennan et al., directed to detection of TNF alpha in synovial cells,
2. Zinner et al., documenting the mRNA level of S100 (a protein that is calcium modulated), and showing that this protein is post-transcriptionally regulated
3. Eriksson et al., which teaches a lack of correlation of mRNA with protein from an insulin responsive glucose transporter
4. Hell et al., on the presence of bcl-2 mRNA in Hodgkin's cells,
5. Guo et al., describing Oatp2 in the liver is regulated at both the transcriptional level and the translational level
6. Fu et al., describing that p53 protein levels may not correlate with p53 mRNA levels
7. Yokota et al., describing retinoblastoma (RB) protein in lung carcinomas. [3]

All of these references are said to disclose a discordance between mRNA expression and protein expression for genes that are *unrelated* to PAGE4.

In contrast to the references cited in the Office action, Orntoft et al., (Molec. Cell Proteomics 1: 37-45, 2002, of record) describes a genome-wide study of gene copy numbers, transcripts and protein levels in pairs of non-invasive and invasive human carcinomas. Although it was only possible to compare mRNA and protein in a few cases (due to a limited ability to focus some of the proteins on two dimensional gels), there was a good correlation ($p<0.005$) between transcript alterations and protein levels. The final Office action asserts that as Orntoft et al. discloses that the expression of seven proteins did not correlate with mRNA levels, the presence of any protein cannot be predicted based on mRNA expression. This is in stark contrast

[3] This publication was cited in the Office action but not listed on a PTO-1449 form and a copy was not provided to the Applicants. A review of the abstract indicates that this reference discloses that absence and shortened mRNA is correlated with the absence of full length RB protein, although in a few cases mRNA was present without protein. Yokota et al. conclude that inactivation of the RB gene (so that protein is not produced) may be important in the development of lung cancer.

to the actual conclusions asserted in Orntoft et al. which states "[i]n general there was a highly significant correlation ($P<0.005$) between mRNA and protein alterations." Orntoft et al. continues to describe a limited set of genes wherein mRNA and protein do not correlate (this is the seven genes referred to on page 11 of the final Office action). However, Orntoft et al. concludes that the lack of correlation between these specific mRNAs and proteins is related to the location of these genes on chromosome 17 (see page 43 of Orntoft et al.). Thus, for other proteins (not belonging to this specific family encoded by chromosome 17) protein expression uniformly correlated to mRNA level. Thus, the results presented in Orntoft et al. support the conclusion that a showing of a correlation between mRNA level and the presence of cancer is sufficient to support the association of the expression of a protein with the presence of cancer.

In addition, the specific data presented in the Declaration of Dr. Pastan negates any allegation that PAGE4 mRNA would not be predictive of the expression and use of PAGE4 protein. The Declaration of Dr. Pastan describes that Northern blot and reverse transcriptase polymerase chain reaction (see the specification, page 4, line 30 to page 5, line 6; page 5, lines 16-20; and FIGS. 3 and 5) were used to evaluate the expression of PAGE4 in prostate cancer. Polyclonal antibodies (see Example 3, page 41 of the specification) were used to demonstrate that PAGE4 protein (SEQ ID NO: 1) was expressed in prostate cancer. Western blot analysis confirmed that PAGE4 protein was expressed in a prostate cancer lysate (see Fig. 2B of Iavrone et al., Mol. Cancer Therap. 1: 329-335, 2002, of record, for an exemplary blot). Samples of prostate cancer from five patients, whose cancers expressed PAGE4 mRNA, were analyzed to confirm that PAGE 4 protein was expressed. PAGE4 protein was expressed in all five of these samples (a 100% correlation). Thus, for PAGE 4, there is a 100% correlation between mRNA expression and protein expression. This specific data with respect to the expression of PAGE4 (SEQ ID NO: 1) negates any general allegations of inoperability based on the prior art describing the expression of unrelated genes. This specific data documents that polypeptides comprising SEQ ID NO: 1 can be used to detect prostate cancer.

The Declaration of Dr. Pastan also describes the synthesis of nine-mer and ten-mer peptides of PAGE4 (SEQ ID NO: 1). Three native PAGE4 peptides of nine to ten consecutive amino acids of PAGE4 (SEQ ID NO: 1) were produced that were predicted to have high binding affinity; two of three native peptides (P16, P84) bound to HLA-A2 molecules with high affinity. Studies were performed to demonstrate that T-cell lines could be generated from peripheral

blood mononuclear cells (PBMC) from prostate cancer patients using polypeptides consisting of 9 to 10 amino acids of SEQ ID NO: 1 that bound MHC class I. These results from these *in vitro* studies confirmed that cytotoxic T cells could be activated (so that they produce a cytokine) using the claimed polypeptides. A single PAGE4 peptide (P16) was then selected for additional study; the results demonstrated that CTLs generated using P16 could lyse prostate tumor cells that express PAGE4 (SEQ ID NO: 1). Thus, the data presented in the Declaration of Dr. Pastan also supports a utility (the production of an immune response to prostate cancer cells) for polypeptides consisting of 8 to 11 amino acids (specifically 9 and 10 amino acids) of SEQ ID NO: 1.

2. Post-filing date evidence can be used to support a disclosed utility

As discussed above, the Declaration of Dr. Pastan documents that samples of prostate cancer from five patients whose cancers express PAGE4 mRNA also express PAGE4 protein (SEQ ID NO: 1). The Declaration of Dr. Pastan also documents that polypeptides consisting of 8 to 11 amino acids (specifically 9 to 10 amino acids) of SEQ ID NO: 1 can be used to produce an immune response to prostate cancer cells. Thus, this Declaration documents that PAGE4 protein (SEQ ID NO: 1) can be used for the detection of cancer.

The Federal Circuit and its predecessor have determined that in those cases where an applicant supplied a reasonable evidentiary showing supporting an asserted therapeutic utility, a 35 U.S.C. 101-based rejection should be reversed (see, for example, *In re Brana*, 51 F.3d 1560, 34 USPQ 1436 (Fed. Cir. 1995); *Cross v. Iizuka*, 753 F.2d 1040, 224 USPQ 739 (Fed. Cir. 1985); *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (CCPA 1980); *Nelson v. Bowler*, 626 F.2d 853, 856, 206 USPQ 881, 883 (CCPA 1980); *In re Malachowski*, 530 F.2d 1402, 189 USPQ 432 (CCPA 1976); *In re Gaubert*, 530 F.2d 1402, 189 USPQ 432 (CCPA 1975); *In re Gazave*, 379 F.2d 973, 154 USPQ 92 (CCPA 1967); *In re Hartop*, 311 F.2d 249, 135 USPQ 419 (CCPA 1962); *In re Krimmel*, 292 F.2d 948, 130 USPQ 215 (CCPA 1961), and MPEP 2107.03 (III). Thus, these decisions make clear that an post-filing evidentiary showing can be used to support the utility of an invention.

In the present application, the Applicants believe that any assertion of the lack of usefulness of the claimed peptides based the prior art cited in the Office action is rendered moot

by the documentation of the specific, substantial and credible use of the claimed polypeptides to detect and treat cancer prostate cancer confirmed in the Declaration of Dr. Pastan.

3. Cancer treatment is an acceptable utility for claimed compositions

The Office action dated September 13, 2005 acknowledges (see page 14) that the polynucleotide encoding SEQ ID NO: 1 is differentially expressed in prostate and uterine cancers. However, the final Office action dated May 16, 2006 appears to assert that as there is no utility in the absence of data documenting therapeutic use in patients as a "potential role as an object of use-testing" (page 9-10).

In the present application, a therapeutic use has been asserted, namely the use of the polypeptides for the treatment and or detection of cancer, such as prostate cancer (see page 17, line 23 to page 19, line 20). The final Office action cites Boon et al. as showing that even if an immune response is achieved, "therapeutic success remains unpredictable" (see page 13 of the final Office action). However, a showing of a failure of one particular therapy for the treatment of cancer does not negate an asserted therapeutic use for all other therapies. The Applicants only need to document a reasonable correlation between the activity and the asserted use.

Indeed, the Courts have found utility for therapeutic inventions despite the fact that an applicant is at a very early stage in the development of a pharmaceutical product or therapeutic regimen based on a claimed pharmacological or bioactive compound or composition. The Federal Circuit, in *Cross v. Iizuka*, 753 F.2d 1040, 1051, 224 USPQ 739, 747-48 (Fed. Cir. 1985) found that data from *in vitro* testing that showed pharmacological activity could be used to support the utility of the claimed product. Indeed, MPEP § 2107 states

> "As a general matter, evidence of pharmacological or other biological activity of a compound will be relevant to an asserted therapeutic use if there is a reasonable correlation between the activity in question and the asserted utility. *Cross v. Iizuka*, 753 F.2d 1040, 224 USPQ 739 (Fed. Cir. 1985); *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (CCPA 1980); *Nelson v. Bowler*, 626 F.2d 853, 206 USPQ 881 (CCPA 1980). An applicant can establish this reasonable correlation by relying on statistically relevant data documenting the activity of a compound or composition, arguments or reasoning, documentary evidence (e.g., articles in scientific journals), or any combination thereof. The applicant does not have to prove that a correlation exists between a particular activity and an asserted therapeutic use of a compound as a matter of statistical certainty, nor does he or she have to provide actual evidence of success in treating humans where such a utility is asserted. Instead, as the courts have repeatedly held, all that is required is a

> reasonable correlation between the activity and the asserted use. *Nelson v. Bowler*, 626 F.2d 853, 857, 206 USPQ 881, 884 (CCPA 1980)." [emphasis added]

In the present case, there is a reasonable certainty that the claimed polypeptides have utility, as evidenced by Ornoft et al. In addition, the Declaration of Dr. Pastan (discussed above) confirms that the claimed peptides have a utility asserted in the specification. The Office action asserts that as the data provided in the Declaration of Dr. Pastan shows only *in vitro* lysis of prostate cancer cell lines, it cannot be used to support or predict the usefulness of PAGE4 peptides to treat cancer (see highlighted phrase on page 18 of the final Office action).

The Office action discounts results in an art accepted model that would be viewed by one of skill in the art as being predictive of the claimed utility. The final Office action contradicts MPEP § 2107.03, which confirms that *in vitro* data can be used to support an asserted utility:

> If an applicant provides data, whether from *in vitro* assays or animal tests or both, to support an asserted utility, and an explanation of why that data supports the asserted utility, the Office will determine if the data and the explanation would be viewed by one skilled in the art as being reasonably predictive of the asserted utility. See, e.g., *Ex parte Maas*, 9 USPQ2d 1746 (Bd. Pat. App. & Inter. 1987); *Ex parte Balzarini*, 21 USPQ2d 1892 (Bd. Pat. App. & Inter. 1991). Office personnel must be careful to evaluate all factors that might influence the conclusions of a person of ordinary skill in the art as to this question, including the test parameters, choice of animal, relationship of the activity to the particular disorder to be treated, characteristics of the compound or composition, relative significance of the data provided and, most importantly, the explanation offered by the applicant as to why the information provided is believed to support the asserted utility. If the data supplied is consistent with the asserted utility, the Office cannot maintain a rejection under 35 U.S.C. §101.

It is disclosed in the present specification that peptides of 8 to 11 amino acids in length of SEQ ID NO: 1 can be used to induce an immune response to prostate cancer and can be used for cancer treatment. There is adequate reason for one of skill in the art to believe that the claimed polypeptides can be used for therapeutic purposes. This is also evidenced by Visseren et al., which discloses that peptides from a related polypeptide, MAGE-2, can be used to produce an immune response against melanoma cells and can be used from the production of peptide based vaccines.

The final Office action cites to a number of references[4] as providing evidence that cancer therapy can be unreliable. Specifically, the Office action refers to (1) Dexler et al., on the characteristics of Hodgkin and Reed-Sternberg cancer cells in vitro; (2) Embelton et al. on procedures for diagnosis of osteogenic sarcoma using monoclonal antibodies; (3) Hsu on the change of chromosomes in cells in culture; (4) Van Dyke et al. on the loss of chromosome 21 in hematologic disease; (5) Kunkel et al. on the expression of hepatocyte growth factor in glioblasomas in culture; and (6) Smith et al., on cancer and the immune system, describing persistent antigen overload created by the shedding of antigens by a growing tumor. None of these references describe experimental results obtained using any specific polypeptides that bind MHC to induce a cytotoxic T cell response to prostate cancer cells. Indeed, if every finding of an artifact associated with a unique type of cancer cell line is to be viewed as negating any possibility of treating cancer and to negate the effectiveness of any proposed therapeutic protocol, no therapeutic intervention for cancer treatment would ever be patentable based on *in vitro* data.

The negative statements made in the final Office action specifically contradict MPEP 2107, which confirms that even in situations where there are no previously successful treatments (and this most certainly is not the case with prostate cancer), there is no basis for a rejection on the basis of utility. MPEP 2107 describes that prior to the 1980's, there were a number of cases where an asserted use in treating cancer in humans was viewed as "incredible." *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (CCPA 1980); *In re Buting*, 418 F.2d 540, 163 USPQ 689 (CCPA 1969); *Ex parte Stevens*, 16 USPQ2d 1379 (Bd. Pat. App. & Inter. 1990); *Ex parte Busse*, 1 USPQ2d 1908 (Bd. Pat. App. & Inter. 1986); *Ex parte Krepelka*, 231 USPQ 746 (Bd. Pat. App. & Inter. 1986); *Ex parte Jovanovics*, 211 USPQ 907 (Bd. Pat. App. & Inter. 1981). Only in those cases where the applicant was unable to come forward with any relevant evidence to rebut a finding by the Office that the claimed invention was inoperative was a 35 U.S.C. § 101 rejection affirmed by the Court (*In re Buting*, 418 F.2d 540, 543, 163 USPQ 689, 690 (CCPA 1969); record did not establish a credible basis for the assertion that the single class of compounds in question would be useful in treating disparate types of cancers). In all of the other cases the treatment of cancer was viewed to be a specific, substantial and credible use.

[4] Copies of these references were not listed on a PTO-1449 form, and were not provided to the Applicants.

The utility of the claimed peptides is further supported by *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (CCPA 1980), wherein an inventor claimed protection for pharmaceutical compositions for treating leukemia. The active ingredient in the compositions was a structural analog to a known anticancer agent. The applicant provided evidence showing that the claimed analogs had the same general pharmaceutical activity as the known anticancer agents. The Court reversed the Board's finding that the asserted pharmaceutical utility was "incredible," pointing to the evidence that showed the relevant pharmacological activity. In the present application, the active ingredient in therapeutic compositions is peptides that bind MHC, and induce a T cell response (the same pharmaceutical activity as known peptides, such as MAGE polypeptides, see Visseren et al.). Thus, one of skill in the art would predict that the therapeutic utility of the claimed peptides is credible.

One of skill in the art has a reasonable basis to believe that the presently claimed polypeptides have a specific, substantial and credible utility, as described in the specification. In addition, data has been provided from an *in vitro* assay confirming the specific, substantial and credible use. As confirmed in MPEP 2107.03, it is inappropriate for the Office to maintain the rejection under 35 U.S.C. § 101.

(3) Claims 1-8, 14-15, 17-18 and 54-57 are fully enabled by the specification.

The final Office action alleges that as claims 1-8, 14-15, 17-18 and 53-57 are not supported by a specific, substantial and credible utility under 35 U.S.C. § 101, because one skilled in the art would not know how to use the claimed invention (see page 19 of the final Office action). Applicants respectfully disagree.

As discussed above, there is a specific, substantial and credible utility for the claimed polypeptides. In determining whether a patent application is in compliance with the enablement requirement, it must be considered whether one of ordinary skill in the art could practice the claimed invention without undue experimentation (*In re Wands*, 858 F.2d 731, 8 USPQ 1400 (Fed. Cir. 1988)). In the present application the claims are directed to (a) polypeptides comprising SEQ ID NO: 1 (claims 1, 2, 7, 14 and 53), (b) polypeptides consisting of 8 to 11 amino acids of SEQ ID NO: 1 (or 9 to 10 amino acids of SEQ ID NO: 1) (claims 2, 4, 6, 8 and 54-57), (c) polypeptides consisting of SEQ ID NO: 1, and (d) the use of these polypeptides

(claims 14 and 15). The final Office action asserts that the claims are directed to polypeptides comprising 8 to 11 amino acids of SEQ ID NO: 1 (see page 20 of the final Office action).

a. Claims 1, 2, 7 and 53 as directed to polypeptides comprising SEQ ID NO:1.

These claims are directed to polypeptides comprising SEQ ID NO: 1 and immunogenic compositions including these proteins. The final Office action does not appear to address this rejection as applied to these claims. The final Office action alleges that the pending claims are not limited to a single amino acid sequence of 102 amino acids, but appears to only address polypeptides comprising 8 to 11 amino acids of SEQ ID NO: 1. The final Office action states "due to the open language of "polypeptide comprising 8 to 11 contiguous amino acids of SEQ ID NO: 1" in claim 1, the polypeptides of claims 1, 2, 7, and 17 encompass unknown sequences attached to 8 to 11 contiguous amino acids of SEQ ID NO: 1" (see the second paragraph on page 20 of the final Office action). The final Office action also asserts that claim 2 encompasses peptides that "not only could bind to MHC I molecule, but also could induce CTL response, in view of the disclosure of the specification" (see the third paragraph on page 20 of the final Office action).

Claims 2 and 7 are directed to polypeptides comprising SEQ ID NO: 1 (not to fragments of SEQ ID NO: 1). Claims 2 and 7 are also directed to an immunogenic composition. Applicants agree with the comments in the Office action that immunogenic compositions of claims 2 and 7 could (1) bind MHC class I, and (2) induce a CTL response, as described in the specification.

Applicants submit that these claims 2, 7 and 53 are fully enabled by the specification, as evidenced by the following Wands analysis.

1. The breadth of the claims

The claims are limited to polypeptides comprising the amino acid sequence set forth as SEQ ID NO: 1.

2. The nature of the invention

The invention is limited to polypeptides including the amino acid sequence set forth as SEQ ID NO: 1, and immunogenic compositions including these polypeptides. Methods for

synthesizing amino acid sequences are routine and/or automated. The final Office action concedes that methods for synthesizing polypeptide sequences are routine and automated (see the final Office action, page 21, second to last line).

3. The state of the prior art

A protein such as PAGE4 (SEQ ID NO: 1) can be chemically synthesized by standard methods. Polypeptides can also be produced using molecular genetic techniques, such as by inserting a nucleic acid encoding SEQ ID NO: 1 or an epitope thereof into an expression vector, introducing the expression vector into a host cell, and isolating the polypeptide. Molecular techniques to produce and isolate proteins are well known in the art. These techniques are described in standard textbooks, for example, *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

4. The level of skill of one of ordinary skill in the art

The level of skill of the average molecular biologist is high.

5. The level of predictability in the art

The production of a protein with a specified sequence is well known. Fusion proteins including PAGE4 (SEQ ID NO:1) can readily be produced (see, for example, the specification at page 34, lines 7-16). Pharmaceutical formulations for immunogenic compositions are well known in the art, see the specification at page 33, line 10 to page 35, line 18. The administration of polypeptides as pharmaceutical compositions is well known in the art; the specification refers to standard texts and U.S. patents that describe formulations for immunogenic compositions including polypeptides (see, for example, page 33, lines 10-14; page 33, line 30-31; page 34, lines 13-16 and 20-24 and page 37, line 33 to page 38, line 34).

6. The amount of direction provided in the application

There is considerable direction provided in the application. The amino acid sequence of PAGE4 is provided as SEQ ID NO: 1 of the specification. This amino acid sequence is 102 amino acids in length. The conjugation of PAGE4 to other agents, such as a lipid is described in

the specification, such as on page 23, lines 21-21. Immunogenic compositions are described in detail in the specification, for example on page 33, line 10 to page 35, line 19 and page 37, line 33 to page 38, line 34.

7. The existence of working examples

SEQ ID NO: 1 is provided in the specification. The isolation of the nucleic acid encoding SEQ ID NO: 1 and the production of expression vectors encoding SEQ ID NO: 1 is described in the examples section of the specification. The Declaration of Dr. Pastan describes the production of immunogenic compositions.

8. The quantity of experimentation needed to make or use the invention

Synthetic polypeptides can be produced by automated machinery. In addition, a number of expression vectors are known in the art (and commercially available) that can be used to produce the claimed polypeptides. Thus, only very limited routine experimentation is required to produce polypeptides comprising SEQ ID NO: 1, and immunogenic compositions including these polypeptides. In addition, the production of pharmaceutical compositions including polypeptides is routine.

Based on the analysis above, it is clear that claims 1, 2, 7 and 53 directed to polypeptides comprising SEQ ID NO:1 are fully enabled by the specification.

b. Claims 2, 4, 6, 8 and 55--57 directed to polypeptides consisting of 8 to 11 amino acids of SEQ ID NO: 1

The final Office action admits that computer programs were available that accurately predicted which polypeptide fragments of a full-length amino acid sequence will bind MHC. However, it appears that the Office action alleges that claims directed to these polypeptide fragments cannot be enabled as it is not clear which of these polypeptide fragments will activate CTLs, and be of use for treatment. The Office action appears to rely on "the unpredictability of protein level," while acknowledging that data documenting changes in the expression of PAGE4 has been presented in the specification (see page 23 of the final Office action).

In the present application, the novel amino acid sequence set forth as SEQ ID NO: 1 is provided, and polypeptides are disclosed that are 8 to 11 amino acids of SEQ ID NO: 1. The specification describes specific functional properties of the claimed peptides of use, namely that *they bind MHC.* As discussed above, polypeptides that bind MHC are of use to induce an immune response. *The pending claims are limited to only those polypeptides that consist of 8 to 11 amino acids of SEQ ID NO: 1 that bind MHC, and compositions including these polypeptides.* The final Office action acknowledges that computer programs are available wherein a technician can type in the amino acid sequence and the program will predict which peptides of a specified length bind MHC.

In this context, Applicants provide the following Wands analysis:

1. The breadth of the claims

The claims are limited to polypeptides consisting of eight to eleven amino acids of SEQ ID NO: 1 that bind MHC (claims 2, 4, 6, 8 and 54-57). Thus, the scope of the claims is limited to fragments of a specified length of a single 102 amino acid sequence that can bind MHC class I. These fragments must include contiguous amino acids.

2. The nature of the invention

The invention is limited to polypeptides that consist of 8 to 11 (or 9 to 10) consecutive amino acids of SEQ ID NO: 1 that bind MHC class I, and to immunogenic compositions including these polypeptides.

3. The state of the prior art

The prior art teaches how to identify immunogenic epitopes of a specified protein sequence that will bind MHC and induce an immune response. Computer programs were available at the time the application was filed wherein a technician can enter a specified amino acid sequence and the computer will predict which amino acid segment will bind MHC, such as HLA-A2.

To identify polypeptides consisting of eight to ten amino acids in length that bind MHC, the PAGE4 amino acid sequence can be entered into a computer program to identify epitopes of interest (see the specification, page 20, line 20 to page 21, line 12). Programs were publicly

available for the identification of epitopes that bind MHC (see Parker et al., Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains, J Immunol. 152:163-75, 1994, provided on the internet at http://bimas.dcrt.nih.gov/molbio/hla_bind/ (print-out enclosed); see also Rammensee et al., Immunogenetics 50: 213-219, 1999 and additional related publications cited in the specification on page 20, line 20 to page 21, line 2). As discussed above, these methods for predicting MHC binding were well known to those of skill in the art at the time the provisional application was filed (see for example, Parker et al., *supra*).

The Office action admits that these methods are known in the art to produce polypeptides (see the final Office action, page 21, second to last line). The final Office action also admits that computer programs were available that would predict which amino acids will bind MHC (see page 22 of the final Office action, middle of the page).

Pharmaceutical formulations for immunogenic compositions are well known in the art, see the specification at page 33, line 10 to page 35, line 18. The administration of polypeptides as pharmaceutical compositions is well known in the art; the specification refers to standard texts and U.S. patents that describe formulations for immunogenic compositions including polypeptides (see, for example, page 33, lines 10-14; page 33, line 30-31; page 34, lines 13-16 and 20-24 and page 37, line 33 to page 38, line 34).

4. The level of skill of one of ordinary skill in the art

The level of skill of the average molecular biologist is high.

5. The level of predictability in the art

Computer programs can be used to predict which eight to ten consecutive amino acids of a specified polypeptide are likely to bind MHC. These programs rank polypeptides in order of predicted strength of the binding. Once the polypeptides are identified, a biological assay can be used to confirm that the eight to ten consecutive amino acids actually bind MHC. The generation of peptide specific cells is known in the art (see Tsang et al., J Natl Cancer Inst 87:982-90, 1995, copy enclosed). Immunogenic compositions can readily be prepared, as discussed above.

6. The amount of direction provided in the application

There is considerable direction provided in the application. The amino acid sequence of PAGE4 is provided as SEQ ID NO: 1 of the specification. This amino acid sequence is 102 amino acids in length. Immunogenic peptides are clearly described in the specification. For example, immunogenic peptides, such as peptides that bind MHC are disclosed in the specification on page 7, line 35 to page 8, line 25, and on page 20, line 1 to page 22, line 5. The specification also discloses that epitopes of use are 8-10 amino acids in length and have anchoring residues. Specific configurations of use are disclosed, such as wherein the PAGE4 polypeptides is 9 or 10 amino acids in length and includes binding motifs for HLA-A2 (see, for example, page 8, lines 30-37, page 20, to page 21, line 2, and page 21, lines 15-19), such as those peptides that have specific anchoring residues in the second position and a positively charged amino acid at the position nine (see page 20, line 20 to page 21, line 2). The selection of binding motifs that bind HLA-A2 is further described on page 28, line 25 to page 29, line 29. Methods and computer based programs for predicting MHC binding motifs (immunogenic epitopes) were disclosed in the specification (for example, see page 8, lines 12-25 and page 21, line 34 to page 22, line 5), and were well known to those of skill in the art at the time the provisional application was filed (see for example, Parker et al., J Immunol 152:163-75, 1994). In addition, biological methods of testing whether a specific epitope is immunogenic are also provided (for example, see page 8, lines 1-4 and page 21, lines 3-12 and lines 20-29).

Pharmaceutical formulations for immunogenic compositions are disclosed, see the specification at page 33, line 10 to page 35, line 18. The specification refers to standard texts and U.S. patents that describe formulations for immunogenic compositions including polypeptides (see, for example, page 33, lines 10-14; page 33, line 30-31; page 34, lines 13-16 and 20-24 and page 37, line 33 to page 38, line 34).

7. The existence of working examples

SEQ ID NO: 1 is provided in the specification, as are experimental examples describing the isolation of the nucleic acid encoding SEQ ID NO: 1 and expression vectors encoding SEQ ID NO: 1. A fragment of PAGE4 (15 amino acids) that can be used to produce antibodies is disclosed on page 41.

In re Bundy, 642 F.2d 430, 434, 209 USPQ 48, 51-52 (CCPA 1981) held that appellant's disclosure was sufficient to enable one skilled in the art to use the claimed analogs of naturally occurring prostaglandins even though the specification lacked any examples of specific dosages, because the specification taught that the novel prostaglandins had certain pharmacological properties and possessed activity similar to known E-type prostaglandins. This is similar to the present application, wherein the specification teaches that the novel peptides have specific pharmacological properties and possess a specified activity (the binding of MHC). Moreover, specific physical properties of the claimed polypeptides are disclosed (the presence of anchor residues).

The Declaration of Dr. Pastan (discussed above in detail) discloses the production of several polypeptides that consist of 9 to 10 amino acids of SEQ ID NO: 1 and bind MHC class I. Immunogenic compositions including these polypeptides were produced and used to induce an immune response.

8. The quantity of experimentation needed to make or use the invention

Once a polypeptide consisting of eight to ten consecutive amino acids of SEQ ID NO: 1 is identified using am art-recognized program, the polypeptide must be synthetically produced. As noted above, the synthesis of short polypeptides is routine. Synthetic polypeptides can be produced by automated machinery. In addition, a number of expression vectors are known in the art (and commercially available) that can be used to produce the claimed polypeptides. Thus, only very limited routine experimentation is required to produce polypeptides consisting of SEQ ID NO: 1 or fragments of 8 to 11 amino acids of SEQ ID NO: 1.

Following synthesis, the polypeptide is screened to demonstrate it binds MHC and can induce a T cell response. Biological methods to test whether a specific epitope is immunogenic are provided in the specification (for example, see page 8, lines 1-4 and page 21, lines 3-12 and lines 20-29). One of skill in the art can perform these assays, as evidenced in the Declaration of Dr. Pastan, which documents the production of the claimed polypeptides and the production of immunogenic compositions including these polypeptides using the guidance provided by the specification.

Thus, given the very complete disclosure provided by the specification, only limited experimentation is required.

c. Claim 54 directed to a polypeptide consisting of SEQ ID NO: 1

The final Office action does not appear to address claim 54, which is directed to a polypeptide consisting of SEQ ID NO: 1. As SEQ ID NO:1 is provided by the specification, applicants believe that claim 54 is fully enabled.

d. Methods for inhibiting the growth of a malignant cell (claims 14-15 and 17-18).

Claims 14-15 and 17-18 directed to the use of PAGE4 to inhibit the growth of a malignant cell are fully enabled by the specification, as evidenced by the following Wands analysis.

1. The breadth of the claims

Claims 14-15 and 17-18 are directed to the use of the claimed polypeptides to inhibit the growth of a malignant cell by activating cytotoxic T lymphocytes.

2. The nature of the invention

The invention is the use of a polypeptide comprising SEQ ID NO: 1 (claim 14 and 17, which depends from claims 53), or a polypeptide consisting of SEQ ID NO: 1 (claim 15 and 18, which depend from claim 54 or a dependent claim thereof) to inhibit the growth of a malignant cell expressing SEQ ID NO: 1. These claims concern activation of the immune system with the specifically claimed polypeptides.

3. The state of the prior art

The use of immunogenic peptides to inhibit the growth of a malignant cell is well known in the art. The state of the art is evidenced by Visseren et al., who describes the use of MAGE proteins for the treatment of melanoma.

4. The level of skill of one of ordinary skill in the art

The level of skill of the average immunologist is high.

5. The level of predictability in the art

Once the sequence of an immunogenic peptide is determined, the use of that peptide to induce an immune response against a cell expressing SEQ ID NO: 1 is predictable. Once the amino acid sequence of a polypeptide is known, that polypeptide can be used to produce activated T cells (Tsang et al., J Natl. Cancer Inst 87:982-90, 1995) that can lyse tumor cells expressing the full-length polypeptide.

6. The amount of direction provided in the application

There is considerable direction provided in the application. The use of PAGE 4 peptides to produce an immune response, and for the treatment of cancer is disclosed in the specification on page 22, line 8 to page 25, line 10.

Pharmaceutical formulations for immunogenic compositions are well known in the art, and are described in the specification, see for example, page 33, line 10 to page 35, line 18. The administration of polypeptides as pharmaceutical compositions is well known in the art; the specification refers to standard texts and U.S. patents that describe formulations for immunogenic compositions including polypeptides (see, for example, page 33, lines 10-14; page 33, line 30-31; page 34, lines 13-16 and 20-24 and page 37, line 33 to page 38, line 34).

7. The existence of working examples

SEQ ID NO: 1 is provided in the specification, as are experimental examples describing the isolation of the nucleic acid encoding SEQ ID NO: 1 and expression vectors encoding SEQ ID NO: 1. The Declaration of Dr. Pastan documents that, using the guidance provided by the specification, cytotoxic lymphocytes (CLTs) could be produced that lyse malignant cells expressing SEQ ID NO: 1.

8. The quantity of experimentation needed to make or use the invention

A polypeptide comprising the amino acid sequence set forth as SEQ ID NO: 1, or consisting of 8 to 11 amino acids of SEQ ID NO: 1, can readily be produced. Following synthesis, an immunogenic polypeptide, such as a polypeptide consisting of 8 to 11 amino acids of SEQ ID NO: 1 can be screened to demonstrate it binds MHC and can induce a T cell response. Biological methods to test whether a specific epitope is immunogenic are provided in the

specification (for example, see page 8, lines 1-4 and page 21, lines 3-12 and lines 20-29), and are routine for one of skill in the art.

The production of antigen specific cytotoxic T cells is routine for one of skill in the art (see Tsang et al., J Natl Cancer Inst 87:982-90, 1995). One of skill in the art can perform these assays. The Declaration of Dr. Pastan provides further evidence that, using methods such as those described in the specification, a polypeptide consisting of nine consecutive amino acids of SEQ ID NO: 1 can be used to activate cytotoxic T cells. These T cells can lyse prostate cancer cells *in vitro*. This evidence supports the assertion that the specification is fully enabling for the use of the claimed polypeptides.

In view of the above remarks, and the Declaration of Dr. Pastan, Applicants believe that they have overcome the rejections under 35 U.S.C. § 112, first paragraph. Applicants request that the rejection of claims 1-8, 14-15, 17-18 and 54-57 be withdrawn.

Respectfully submitted,

KLARQUIST SPARKMAN, LLP

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 595-5300
Facsimile: (503) 595-5301

By _____
Susan Alpert Siegel, Ph.D.
Registration No. 43,121

Appendix

Listing of Claims Subject to Appeal

1. (Rejected) An isolated polypeptide comprising:

(a) an amino acid sequence set forth as SEQ ID NO: 1; or

(b) 8 to 11 contiguous amino acids of SEQ ID NO: 1, wherein the polypeptide binds major histocompatibility complex (MHC) I.

2. (Rejected) An immunogenic composition comprising the isolated polypeptide of claim 53, and a pharmaceutically acceptable carrier.

3. (Canceled)

4. (Rejected) An immunogenic composition comprising the isolated polypeptide of claim 54, and a pharmaceutically acceptable carrier.

5. (Canceled)

6. (Rejected) The isolated polypeptide of claim 54, wherein the isolated polypeptide is conjugated to a lipid.

7. (Rejected) The immunogenic composition of claim 2, further comprising two or more of a stabilizing detergent, a micelle-forming agent, and an oil.

8. (Rejected) The immunogenic composition of claim 4, further comprising two or more of a stabilizing detergent, a micelle-forming agent, and an oil.

9-12. (Withdrawn)

13. (Canceled)

14. (Rejected) A method for inhibiting the growth of a malignant cell expressing SEQ ID NO: 1, comprising:

(a) culturing cytotoxic T lymphocytes (CTLs) or CTL precursor cells with the polypeptide of claim 53, thus activating the CTLs or CTL precursors; and

(b) contacting the malignant cell with the activated CTLs or CTLs matured from the CTL precursors, thereby inhibiting the growth of the malignant cell.

15. (Rejected) A method for inhibiting the growth of a malignant cell expressing SEQ ID NO: 1 in a mammal with a malignancy, the method comprising:

(a) obtaining cytotoxic T lymphocytes (CTLs) or CTL precursor cells from the mammal;

(b) culturing the CTLs or CTL precursors with the polypeptide of claim 54, thus activating the CTLs or CTL precursors; and

(c) introducing the activated CTLs or CTL precursors into the mammal, thereby inhibiting the growth of the malignant cell.

16. (Withdrawn)

17. (Rejected) A method for inhibiting the growth of a malignant cell expressing SEQ ID NO: 1 in a mammal with a malignancy, the method comprising administering to the mammal a therapeutically effective amount of the immunogenic composition of claim 2, thereby inhibiting the growth of the malignant cell.

18. (Rejected) A method for inhibiting the growth of a malignant cell expressing SEQ ID NO: 1 in a mammal with a malignancy, the method comprising administering to the mammal a therapeutically effective amount of the immunogenic composition of claim 4, thereby inhibiting the growth of the malignant cell.

19-52 (Canceled).

53. (Rejected) The isolated polypeptide of claim 1, comprising an amino acid sequence set forth as SEQ ID NO: 1.

54. (Rejected) An isolated polypeptide consisting of 8 to 11 contiguous amino acids of SEQ ID NO: 1, wherein the polypeptide binds major histocompatibility complex (MHC) I .

55. (Rejected) The isolated polypeptide of claim 54, wherein the polypeptide is 9 to 10 amino acids in length.

56. (Rejected) The isolated polypeptide of claim 53, wherein the polypeptide binds HLA-A1, HLA-A2.1, HLA-A3.2, HLA-A4.1 or HLA-A11.2.

57. (Rejected) The isolated polypeptide of claim 54, conjugated to a lipid.

58-60 (Withdrawn).

Evidence Appendix

1. Declaration of Dr. Pastan under 37 C.F.R. § 1.132, submitted on March 13, 2006.

2. Figure 4-1 of Abbas et al., "Cellular and Molecular Immunology," 4th edition, W. B. Saunders Co. Philadelphia, 2000 (1st edition, 1991), page 64.

3. Grey et al., PCT Publication No. WO 94/020127 A1, cited in the Office action dated May 16, 2006.

4. Iavrone et al., *Molecular Cancer Therapeutics*, "PAGE4 is a Cytoplasmic Protein that is Expressed in Normal Prostate and in Prostate Cancers," 1:329-335, 2002, submitted on March 13, 2006.

5. Orntoft et al., *Molecular & Cellular Proteomics*, "Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas, 1:37-45, 2002, submitted on March 13, 2006.

6. Parker et al., *Journal of Immunology*, "Scheme for Ranking Potential HLA-A2 Binding Peptides Based on Independent Binding of Individual Peptide Side-Chains," 152:163-75, 1994.

7. Print-out of http://bimas.dcrt.nih.gov/molbio/hla_bind/

8. Rammensee et al., *Immunogenetics*, "SYFPEITHI: database for MHC ligands and peptide motifs," 50:213-219, 1999.

9. Schirle et al., *J. Immunological Methods*, "Combining computer algorithms with experimental approaches permits the rapid and accurate identification of T cells epitopes from defined antigens," 257:1-16, 2001.

10. Tsang et al., *J Natl Cancer Inst*, "Generation of Human Cytotoxic T Cells Specific for Human Carcinoembryonic Antigen Epitopes From Patients Immunized with Recombinant Vaccinia-CEA Vaccine," 87:982-90, 1995.

11. Visseren et al., *Int. J. Cancer*, "Identification of HLA-A *0201-Restricted CTL Epitopes Encoded by the Tumor-Specific *MAGE*-2 Gene Product," 73:125-30, 1997, cited in the Office action dated May 16, 2006.

Related Proceedings Appendix

There are no related proceedings.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Pastan et al.
Application No. 09/763,393
Filed: July 30, 2001
Confirmation No. 5265
For: PAGE-4, AN X-LINKED GAGE-LIKE GENE EXPRESSED IN NORMAL AND NEOPLASTIC PROSTATE, TESTIS AND UTERUS, AND USES THEREFOR
Examiner: Minh-Tam Davis
Art Unit: 1642
Attorney Reference No. 4239-61541-01

CERTIFICATE OF MAILING

I hereby certify that this paper and the documents referred to as being attached or enclosed herewith are being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: MAIL STOP AMENDMENT COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450 on the date shown below.

Attorney or Agent for Applicant(s)

Date Mailed March 13, 2006

MAIL STOP AMENDMENT
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450

DECLARATION OF DR. PASTAN UNDER 37 CFR § 1.132

1. I, Ira Pastan, M.D., am an inventor of the above-referenced application.

2. It is my understanding that claims 1-8, 14-15, 17-18, 53, 55, 57 were rejected in the Office action dated August 5, 2004, as allegedly not being enabled by the specification. In addition, claims 1-8, 14-15, 17-18 53, 55 and 57 were rejected as allegedly there is insufficient written description for one of skill in the art to make and use the claimed invention, as polypeptides with "unknown structure" are included. In addition, claims 1-8, 14-15, 17-18, 53, 55 and 57 are rejected as not having any utility, as allegedly one of skill in the art would not know that PAGE4 polypeptides, and immunogenic fragments thereof, could be used to treat cancer based on the specification.

3. The specification includes sufficient written description of PAGE4 (shown in SEQ ID NO: 1), and immunogenic epitopes of PAGE4. The amino acid sequence of PAGE4 is provided as SEQ ID NO: 1 of the specification. This amino acid sequence is 102 amino acids in length.

Immunogenic peptides are clearly described in the specification. For example, immunogenic peptides, such as peptides that bind MHC are disclosed in the specification on page 7, line 35 to page 8, line 25, and on page 20, line 1 to page 22, line 5. The specification also discloses that epitopes of use are 8-10 amino acids in length and have anchoring residues. Specific configurations of use are disclosed, such as wherein the PAGE4 polypeptides is 9 or 10 amino acids in length and includes binding motifs for HLA-A2 (see, for example, page 8, lines 30-37, page 20, to page 21, line 2, and page 21, lines 15-19), such as those peptides that have specific anchoring residues in the second position and a positively charged amino acid at the position nine (see page 20, line 20 to page 21, line 2). The selection of binding motifs that bind HLA-A2 is further described on page 28, line 25 to page 29, line 29). Methods and computer based programs for predicting MHC binding motifs (immunogenic epitopes) were disclosed in the specification (for example, see page 8, lines 12-25 and page 21, line 34 to page 22, line 5), were well known to those of skill in the art at the time the provisional application was filed (see for example, Parker et al., Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152:163-75, 1994). In addition, biological methods of testing whether a specific epitope is immunogenic are also provided (for example, see page 8, lines 1-4 and page 21, lines 3-12 and lines 20-29).

4. Using the information disclosed in the specification, the primary amino acid sequence of human PAGE4 was analyzed for consensus motifs for novel HLA-A2 binding peptides using a computer program. Specifically, the amino acid sequence of PAGE4 was scanned for matches to consensus motifs for HLA-A2 binding peptides. My colleagues and I used the computer algorithm from the BioInformatics and Molecule Analysis Section of NIH (BIMAS) (developed by Parker et al. (see the article published in J Immunol 152:163-75, 1994 for a complete description), which ranks potential MHC binding peptides according to the predictive one-half-time dissociation of peptide/MHC complexes. The HLA-A2 allele was chosen because it is the most commonly expressed class I allele. Nine-mer and ten-mer peptides were synthesized if they conformed to the respective consensus motif; binding of immunogenic peptides to this allele is disclosed in the specification, such as on page 20, lines 10-19. A panel of PAGE4 peptides (Table 1) were made by Biosynthesis Inc. (Lewisville, Texas) with a purity >95%. A MUC-1 peptide and a carcinoembryonic antigen (CEA) HLA-A3 binding peptide (CAP-7) were used as

a positive and negative control, respectively (Tsang et al, J Natl Cancer Inst 87:982-90, 1995). Binding of PAGE4 peptides and the PAGE4 analogues to HLA-A2 molecules was evaluated by the upregulation of HLA-A2 expression on T2 cells as demonstrated by flow cytometry (Njiman et al., Eur J Immunol 23:1215-9, 1993). As shown in Table 1, two of three native peptides (P16, P84) bound to HLA-A2 molecules in the T2 assay.

Table 1. Binding of PAGE4 peptides to HLA-A2 molecules

Peptides	Designation	Amino acid position	Peptide sequence	T2 binding[a]
P16 (native)	test	16-25	QEAPDVVAFV	309 (1.2)
P59 (native)	test	59-68	VEGDCQEMDL	246 (0.9)
P84 (native)	test	84-92	KTPPNPKHA	270 (1.04)
CAP-7 peptide	(negative control)		HLFGYSWYK	260

Studies were then undertaken to examine the stability of the peptide-MHC complex for the peptides P16 and P84. Each peptide was incubated with T2 cells overnight, the unbound peptides were washed off, and the cells were then incubated with Brefeldin A to block delivery of new class I molecules to the cell surface. Cells were analyzed for the presence of peptide-HLA-A2 complexes at various time points. For both peptides, more than 43.6% of complexes remained over the 8-hour observation period.

5. Studies were conducted to determine whether T-cell lines could be generated from peripheral blood mononuclear cells (PBMC) from prostate cancer patients. Autologous dendritic cells (DCs) were used as antigen presenting cells (APC). PAGE4-specific T-cell lines were generated from a prostate cancer patient (termed patient "A") using P16 and P84. Specifically, modification of the protocol described by Tsang et al. (J Natl Cancer Inst 87:982-90, 1995) was used to generate PAGE4-specific CTLs. Irradiated (3000 rad) autologous DCs were used as APCs.

Autologous non-adherent cells were stimulated in the presence of autologous DCs pulsed with peptides at a concentration of 12.5 µg/ml at an effector-to-APC ratio of 10:1. Cultures were maintained for three initial days in medium containing 10% human AB serum, and four

additional days in the same medium supplemented with 20 units/ml of recombinant human IL-2. After a 7-day culture period, designated as an *in vitro* stimulation (IVS) cycle, cells were restimulated as described above for a total of three IVS. After the third IVS cycle, irradiated (23,000 rads) autologous Epstein-Barr virus (EBV) - transformed B cells were used as APCs. The autologous EBV-transformed B cells were pulsed with 12.5 μg/ml of peptides at an effector-to-APC ratio of 1:3.

The T-cell lines were designated T-A-P16 or T-A-P84. The specificity of the PAGE4-specific T cells was analyzed for their ability to release IFN-γ after stimulation with autologous B cells pulsed with the corresponding peptides. Specifically, supernatants of T cells stimulated for 48 hours with peptide-pulsed autologous EBV-transformed B cells, in IL-2-free medium at various peptide concentrations, were screened for secretion of IFN-γ using an ELISA kit (BioSource International, Camarillo, CA) and lymphotactin using an ELISA assay (Muller et al., Eur J Immunol 25: 1744-8, 1995).

As shown in Table 2, high levels of IFN-γ production were observed when the T-cell lines were stimulated with the specific peptide, although higher levels of IFN-γ production was observed for the T-A-P16 T-cell line for the T-A-P84 T-cell line.

Table 2. Production of IFN-γ by T-cell lines generated from a prostate cancer patient stimulated with P16 or P84

	Production of IFN-γ (pg/ml)	
T-cell line	Corresponding peptide	None
T-A-P16	302.0	<15.6
T-A-P84	97.5	<15.6

6. To examine the ability of the P16 peptide to activate the PAGE4-specific T-cells, T-A-P16 cells were analyzed to determine their ability to lyse peptide-pulsed targets. Specifically, a 6-hour or 16-hour ^{111}In release assay was used to determine T-cell mediated killing (Tsang et al., J Natl Cancer Inst 87:982-90, 1995. Target cells were labeled with ^{111}In oxine (Amersham Health, Silver Spring, MD) for 20 minutes at room temperature. For these assays, 3 X 10^3 cells were used per well, in 96-well rounded-bottom culture plates. ^{111}In release was measured by

gamma counting. Spontaneous release was determined by incubating the target cells with medium alone, and complete lysis by incubating the target cells with 2.5% Triton X-100. Specific lysis (%) = [(observed release – spontaneous release)/(complete release – spontaneous release)] X 100.

The results were expressed as percent specific lysis at effector-to-target ratio of 30:1 and 15:1. Labeled T2 cells were incubated with or without peptide (12.5 μg/ml) in serum-free medium for 2 hours at 37°C prior to their addition into the assay. Effector cells were used at IVS4. As shown in Table 3, lysis of T2 cells pulsed with the P16 peptide lysed target cells, at two different E:T cell ratios.

Table 3. Ability of the PAGE4-specific T-cell lines (T-A-P16) to lyse peptide-pulsed targets

Target	% lysis (±SD)			
	T-A-P16		T-B-P16-1	
	30 : 1	15 : 1	30 : 1	15 : 1
T2	5.1 (1.1)	6.4 (0.2)	7.6 (0.1)	6.6 (0.9)
T2 + P16	12.7 (1.7)	9.1 (0.9)	23.5 (0.6)	22.7 (0.8)

CTL degranulation is a requisite process of perforin-granzyme mediated killing by activated CD8+ T cells. Thus, to confirm that the CD8+ cells were indeed activated, CD107a mobilization to the cell surface of CD8+ T cells was examined following activation with P16 peptide in addition to CTL assays. The CD107a mobilization assay was performed according to a method known in the art (see Rubio et al., Nat Med 9:1377-82, 2003). CD8+ T cells stimulated with P16 expressed surface CD107a.

7. Studies were then conducted to determine whether the PAGE4-specific T-cell lines could lyse tumor cells that endogenously express native PAGE4. The expression of HLA-A2 and PAGE4 on tumor cell lines was analyzed by flow cytometry and RT-PCR, respectively. It was demonstrated that T-A-P16 cells were capable of lysing LNCaP human prostate cancer cells that express native PAGE4 and are HLA-A2 positive. At an effector: target ratio of 30:1 approximately 15% of the target cells were lysed, while at an effector:target ratio of 15:1 approximately 8% of the target cells were lysed.

8. These studies demonstrate that immunogenic fragments of PAGE4 can be produced and used to activate lymphocytes, as described in the specification. In addition, the results demonstrate that immunogenic PAGE4 fragments can be used to induce an immune response that results in the lysis of prostate cancer cells. The results presented above (points 2-7) support that one of skill in the art, such as myself, could readily use the disclosure of the above-referenced application to produce peptide fragments of SEQ ID NO: 1 and test their immunogenicity. I and my colleagues have produced variant peptide sequences (with an altered amino acid sequence) that have a high binding affinity for MHC, and can be used to activate T cells. These variant polypeptides are disclosed in a U.S. Provisional Application filed on February 24, 2005 (inventors are Jeffrey Schlom, Kwong-Yok Tsang, and Ira Pastan).

9. It is my understanding that claims 1-8, 14-15, 17-18, 53, 55, 57 were rejected in the Office action dated August 5, 2004, as allegedly one of skill in the art could not determine that SEQ ID NO: 1 or a fragment thereof would have a credible utility. The Office action asserts that one of skill in the art could not predict, based on an mRNA analysis, that encoded protein would be differentially expressed in prostate and uterine cancer. It is my understanding that the Office action asserts (page 14) that protein levels cannot be predictably correlated with either steady state mRNA levels or alterations in mRNA levels for any cancer, thus there cannot be a utility for PAGE4 protein. I disagree with this assertion.

PAGE4 expression is highly expressed in prostate and uterine cancer. Orntoft et al. (Molec. Cell Proteomics 1: 37-45, 2002, copy submitted herewith as Exhibit A) has performed a genome-wide study of gene copy numbers, transcripts and protein levels in pairs of non-invasive and invasive human carcinomas. Although it was only possible to compare mRNA and protein in a few cases (due to a limited ability to focus some of the proteins on two dimensional gels), there was a good correlation ($p<0.005$) between transcript alterations and protein levels. Based on this study, one of skill in the art would predict that the presence (or absence) of PAGE4 mRNA would correlate with the presence (or absence of PAGE4 protein).

Northern blot and reverse transcriptase polymerase chain reaction (see page 4, line 30 to page 5, line 6; page 5, lines 16-20; and FIGS. 3 and 5) were used to evaluate the expression of PAGE4 in cancer. Polyclonal antibodies (see Example 3, page 41 of the specification) were used

to determine if PAGE4 protein was expressed in prostate cancer. Western blot analysis confirmed that PAGE4 was expressed in a prostate cancer lysate (see Fig. 2B of Iavrone et al., Mol. Cancer Therap. 1: 329-335, 2002, copy submitted herewith as Exhibit B). We analyzed samples of prostate cancer from five patients whose cancers expressed PAGE4 mRNA. PAGE4 protein was expressed in all five of these samples (a 100% correlation). To determine the localization of PAGE4 in the cell, nuclear cytoplasmic and membrane fractions (see Wolfgang et al., PNAS 97: 9437-9443, 2000) were prepared from NIH3T3 cells and prostate cancer cells (PC3) stably expressing PAGE4 mRNA. The PAGE4 protein product was detected in the cytoplasmic fraction of these cell lines. Immunohistochemical analyses were performed on NIH3T3 cells transfected with an expression vector encoding PAGE 4. These immunohistochemical studies confirmed that the PAGE4 protein can be detected in the cytoplasm (see FIG. 4 of Iavrone et al., *supra*). The correlation of PAGE4 mRNA with the presence of PAGE4 protein in prostate cancer, and the showing that PAGE4 protein can be detected in cells expressing PAGE4 mRNA documents the utility of the PAGE4 protein.

10. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of the Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Ira Pastan, M.D.

3/10/06
Date

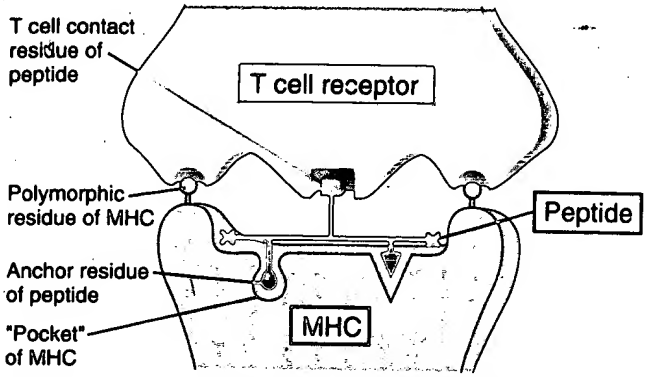


Figure 4–1 T cell recognition of a peptide-MHC complex. This schematic illustration shows an MHC molecule binding and displaying a peptide and a T cell receptor recognizing two polymorphic residues of the MHC molecule and one residue of the peptide. Details of the interactions among peptides, MHC molecules, and T cell receptors are described in Chapters 4, 5, and 6. MHC, major histocompatibility complex.

DISCOVERY OF THE MAJOR HISTOCOMPATIBILITY COMPLEX AND ITS ROLE IN IMMUNE RESPONSES

Discovery of the Mouse Major Histocompatibility Complex

The MHC was discovered as the genetic locus whose products were responsible for rapid rejection of tissue grafts exchanged between inbred strains of mice. Key to understanding this discovery is the concept of genetic polymorphism. Some genes are represented by only one normal nucleic acid sequence in all the members of an animal population (except for relatively rare mutations); such genes are said to be nonpolymorphic, and the normal, or wild-type, gene sequence is usually present on both chromosomes of a pair in every member of the species. (Recall that all chromosomes, except sex chromosomes in the male, are present in pairs in a normal diploid animal.) In the case of other genes, alternative forms, or variants, are present at stable frequencies in different members of the population. Such genes are said to be **polymorphic,** and each common variant of a polymorphic gene is called an **allele.** For polymorphic genes, an individual can have the same allele at that genetic locus on both chromosomes of the pair and is said to be homozygous, or an individual can have two different alleles, one on each chromosome, and is termed heterozygous.

In the 1940s, George Snell and his colleagues used genetic techniques to analyze the rejection of transplanted tumors and other tissues grafted between strains of laboratory mice. To do this, it was necessary to first produce inbred strains by repetitive mating of siblings. After about 20 generations, every mouse of an inbred strain has identical nucleic acid sequences at all locations on all chromosomes. In other words, inbred mice are homozygous at every genetic locus and every mouse of an inbred strain is genetically identical (**syngeneic**) to every other mouse of the same strain. In the case of polymorphic genes, each inbred strain, because it is homozygous, expresses a single allele from the original population. Different strains may express different alleles and are said to be **allogeneic** to one another.

When a tissue or an organ, such as a patch of skin, is grafted from one animal to another, two possible outcomes may ensue. In some cases, the grafted skin survives and functions as normal skin. In other cases, the immune system destroys the graft, a process called **rejection.** Skin-grafting experiments showed that grafts exchanged between animals of one inbred strain are accepted, whereas grafts exchanged between animals of different inbred strains (or between outbred animals) are rejected (Fig. 4–2). Therefore, the recognition of a graft as self or foreign is an inherited trait. The genes responsible for causing a grafted tissue to be perceived as similar to or different from one's own tissues were called **histocompatibility genes,** and the differences between self and foreign were attributed to genetic polymorphisms among different histocompatibility gene alleles.

The tools of genetics, namely breeding and analysis of the offspring, were then applied to identify the relevant genes (Box 4–1). The critical strategy in this effort was the breeding of congenic mouse strains; in two congenic strains, the mice are identical at all loci except the one at which they are selected to be different. Analyses of congenic mice that were selected for their ability to reject grafts from one another indicated that a single genetic region is primarily responsible for rapid graft rejection, and this region was called the major histocompatibility locus. The particular locus that was identified in mice by Snell's group was linked to a gene on chromosome 17 encoding a polymorphic blood group antigen called Antigen II, and therefore this region was named histocompatibility-2 or, simply, H-2. Initially, this locus was thought to contain a single gene that controlled tissue compatibility. However, occasional recombination events occurred within the H-2 locus during interbreeding of different strains, indicating that it actually contained several different but closely linked genes, each involved in graft rejection. The genetic region that controlled graft rejection and contained several linked genes was named the major histocompatibility complex, or MHC. Genes that determine the fate of grafted tissues are present in all mammalian species, are homologous to the *H-2* genes first identified in mice, and are all called MHC genes (Fig. 4–3). Other genes that contribute to graft rejection to a lesser degree are called minor histocompatibility genes; we will return to these in Chapter 16, when we discuss transplantation immunology. The nomenclature of mouse MHC genes is described in Box 4–1.

For almost 20 years after the MHC was discovered, its only documented role was in graft rejection. This was a puzzle to immunologists, because transplantation is not a normal phenomenon, and there was no obvious reason why a set of genes should be

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 September 1994 (15.09.1994)

PCT

(10) International Publication Number
WO 94/020127 A1

(51) International Patent Classification[5]: A61K 37/02, C07K 7/06

(21) International Application Number: PCT/US94/02353

(22) International Filing Date: 4 March 1994 (04.03.1994)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

08/027,146	5 March 1993 (05.03.1993)	US
08/073,205	4 June 1993 (04.06.1993)	US
08/159,184	29 November 1993 (29.11.1993)	US

(71) Applicant: CYTEL CORPORATION [US/US]; 3525 John Hopkins Court, San Diego, CA 92121 (US).

(72) Inventors: GREY, Howard, M.; 9066 La Jolla Shores Lane, La Jolla, CA 92037 (US). SETTE, Alessandro; 5551 Linda Rosa Avenue, La Jolla, CA 92037 (US). SIDNEY, John; 8541 D Villa La Jolla Drive, La Jolla, CA 92037 (US). KAST, W., Martin; Maria Rutgersweg 106, NL-2331 NX Leiden (NL).

(74) Agents: BASTIAN, Kevin, L. et al.; Townsend and Townsend Khourie and Crew, Steuart Street Tower, 20th floor, One Market Plaza, San Francisco, CA 94105 (US).

(81) Designated States *(national)*: AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, LV, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, UA, UZ, VN.

(84) Designated States *(regional)*: European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(48) Date of publication of this corrected version:
17 April 2003

(15) Information about Correction:
see PCT Gazette No. 16/2003 of 17 April 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 94/020127 A1

(54) Title: HLA-A2.1 BINDING PEPTIDES AND THEIR USES

(57) Abstract: The present invention provides the means and methods for selecting immunogenic peptides and the immunogenic peptide compositions capable of specifically binding glycoproteins encoded by HLA-A2.1 allele and inducing T cell activation in T cells restricted by the A2.1 allele. The peptides are useful to elicit an immune response against a desired antigen.

HLA-A2.1 BINDING PEPTIDES AND THEIR USES

The present application is a continuation in part of USSN 08/159,184, which is a continuation in part of USSN 08/073,205, which is a continuation in part of USSN 08/027,146, all of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates to compositions and methods for preventing, treating or diagnosing a number of pathological states such as viral diseases and cancers. In particular, it provides novel peptides capable of binding selected major histocompatibility complex (MHC) molecules and inducing an immune response.

MHC molecules are classified as either Class I or Class II molecules. Class II MHC molecules are expressed primarily on cells involved in initiating and sustaining immune responses, such as T lymphocytes, B lymphocytes, macrophages, etc. Class II MHC molecules are recognized by helper T lymphocytes and induce proliferation of helper T lymphocytes and amplification of the immune response to the particular immunogenic peptide that is displayed. Class I MHC molecules are expressed on almost all nucleated cells and are recognized by cytotoxic T lymphocytes (CTLs), which then destroy the antigen-bearing cells. CTLs are particularly important in tumor rejection and in fighting viral infections.

The CTL recognizes the antigen in the form of a peptide fragment bound to the MHC class I molecules rather than the intact foreign antigen itself. The antigen must normally be endogenously synthesized by the cell, and a portion of the protein antigen is degraded into small peptide fragments in the cytoplasm. Some of these small peptides translocate into a pre-Golgi compartment and interact with class I heavy chains to facilitate proper folding and association with the subunit $\beta 2$ microglobulin. The

peptide-MHC class I complex is then routed to the cell surface for expression and potential recognition by specific CTLs.

Investigations of the crystal structure of the human MHC class I molecule, HLA-A2.1, indicate that a peptide binding groove is created by the folding of the α1 and α2 domains of the class I heavy chain (Bjorkman et al., Nature 329:506 (1987). In these investigations, however, the identity of peptides bound to the groove was not determined.

Buus et al., Science 242:1065 (1988) first described a method for acid elution of bound peptides from MHC. Subsequently, Rammensee and his coworkers (Falk et al., Nature 351:290 (1991) have developed an approach to characterize naturally processed peptides bound to class I molecules. Other investigators have successfully achieved direct amino acid sequencing of the more abundant peptides in various HPLC fractions by conventional automated sequencing of peptides eluted from class I molecules of the B type (Jardetzky, et al., Nature 353:326 (1991) and of the A2.1 type by mass spectrometry (Hunt, et al., Science 225:1261 (1992). A review of the characterization of naturally processed peptides in MHC Class I has been presented by Rötzschke and Falk (Rötzschke and Falk, Immunol. Today 12:447 (1991).

Sette et al., Proc. Natl. Acad. Sci. USA 86:3296 (1989) showed that MHC allele specific motifs could be used to predict MHC binding capacity. Schaeffer et al., Proc. Natl. Acad. Sci. USA 86:4649 (1989) showed that MHC binding was related to immunogenicity. Several authors (De Bruijn et al., Eur. J. Immunol., 21:2963-2970 (1991); Pamer et al., 991 Nature 353:852-955 (1991)) have provided preliminary evidence that class I binding motifs can be applied to the identification of potential immunogenic peptides in animal models. Class I motifs specific for a number of human alleles of a given class I isotype have yet to be described. It is desirable that the combined frequencies of these different alleles should be high enough to cover a large fraction or perhaps the majority of the human outbred population.

Despite the developments in the art, the prior art has yet to provide a useful human peptide-based vaccine or

therapeutic agent based on this work. The present invention provides these and other advantages.

SUMMARY OF THE INVENTION

The present invention provides compositions comprising immunogenic peptides having binding motifs for HLA-A2.1 molecules. The immunogenic peptides, which bind to the appropriate MHC allele, are preferably 9 to 10 residues in length and comprise conserved residues at certain positions such as positions 2 and 9. Moreover, the peptides do not comprise negative binding residues as defined herein at other positions such as positions 1, 3, 6 and/or 7 in the case of peptides 9 amino acids in length and positions 1, 3, 4, 5, 7, 8 and/or 9 in the case of peptides 10 amino acids in length. The present invention defines positions within a motif enabling the selection of peptides which will bind efficiently to HLA A2.1.

Epitopes on a number of immunogenic target proteins can be identified using the peptides of the invention. Examples of suitable antigens include prostate cancer specific antigen (PSA), hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. The peptides are thus useful in pharmaceutical compositions for both in vivo and ex vivo therapeutic and diagnostic applications.

Definitions

The term "peptide" is used interchangeably with "oligopeptide" in the present specification to designate a series of residues, typically L-amino acids, connected one to the other typically by peptide bonds between the alpha-amino and carbonyl groups of adjacent amino acids. The oligopeptides of the invention are less than about 15 residues in length and usually consist of between about 8 and about 11 residues, preferably 9 or 10 residues.

An "immunogenic peptide" is a peptide which comprises an allele-specific motif such that the peptide will

bind an MHC molecule and induce a CTL response. Immunogenic peptides of the invention are capable of binding to an appropriate HLA-A2.1 molecule and inducing a cytotoxic T cell response against the antigen from which the immunogenic peptide is derived.

Immunogenic peptides are conveniently identified using the algorithms of the invention. The algorithms are mathematical procedures that produce a score which enables the selection of immunogenic peptides. Typically one uses the algorithmic score with a "binding threshold" to enable selection of peptides that have a high probability of binding at a certain affinity and will in turn be immunogenic. The algorithm is based upon either the effects on MHC binding of a particular amino acid at a particular position of a peptide or the effects on binding of a particular substitution in a motif containing peptide.

A "conserved residue" is an amino acid which occurs in a significantly higher frequency than would be expected by random distribution at a particular position in a peptide. Typically a conserved residue is one where the MHC structure may provide a contact point with the immunogenic peptide. One to three, preferably two, conserved residues within a peptide of defined length defines a motif for an immunogenic peptide. These residues are typically in close contact with the peptide binding groove, with their side chains buried in specific pockets of the groove itself. Typically, an immunogenic peptide will comprise up to three conserved residues, more usually two conserved residues.

As used herein, "negative binding residues" are amino acids which if present at certain positions (for example, positions 1, 3 and/or 7 of a 9-mer) will result in a peptide being a nonbinder or poor binder and in turn fail to be immunogenic i.e. induce a CTL response.

The term "motif" refers to the pattern of residues in a peptide of defined length, usually about 8 to about 11 amino acids, which is recognized by a particular MHC allele. The peptide motifs are typically different for each human MHC

allele and differ in the pattern of the highly conserved residues and negative residues.

The binding motif for an allele can be defined with increasing degrees of precision. In one case, all of the conserved residues are present in the correct positions in a peptide and there are no negative residues in positions 1,3 and/or 7.

The phrases "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. Thus, the peptides of this invention do not contain materials normally associated with their in situ environment, e.g., MHC I molecules on antigen presenting cells. Even where a protein has been isolated to a homogenous or dominant band, there are trace contaminants in the range of 5-10% of native protein which co-purify with the desired protein. Isolated peptides of this invention do not contain such endogenous co-purified protein.

The term "residue" refers to an amino acid or amino acid mimetic incorporated in an oligopeptide by an amide bond or amide bond mimetic.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a flow diagram of an HLA-A purification scheme.

Fig. 2 shows a scattergram of the log of relative binding plotted against the "Grouped Ratio" algorithm for 9 mer peptides.

Fig. 3 shows a scattergram of the log of relative binding plotted against the average "Log of Binding" algorithm score for 9 mer peptides.

Figs. 4 and 5 show scattergrams of a set of 10-mer peptides containing preferred residues in positions 2 and 10 as scored by the "Grouped Ratio" and "Log of Binding" algorithms.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention relates to the determination of allele-specific peptide motifs for human Class I MHC (sometimes referred to as HLA) allele subtypes, in particular, peptide motifs recognized by HLA-A2.1 alleles. These motifs are then used to define T cell epitopes from any desired antigen, particularly those associated with human viral diseases, cancers or autoiummune diseases, for which the amino acid sequence of the potential antigen or autoantigen targets is known.

Epitopes on a number of potential target proteins can be identified in this manner. Examples of suitable antigens include prostate specific antigen (PSA), hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, melanoma antigens (e.g., MAGE-1), human immunodeficiency virus (HIV) antigens and human papilloma virus (HPV) antigens.

The peptides of the invention may also be employed to relieve the symptoms of, treat or prevent the occurrence or reoccurrence of autoimuune diseases. Such diseases include, for example, multiple sclerosis (MS), rheumatoid arthritis (RA), Sjogren syndrome, scleroderma, polymyositis, dermatomyositis, systemic lupus erythematosus, juvenile rheumatoid arthritis, ankylosing spondylitis, myasthenia gravis (MG), bullous pemphigoid (antibodies to basement membrane at dermal-epidermal junction), pemphigus (antibodies to mucopolysaccharide protein complex or intracellular cement substance), glomerulonephritis (antibodies to glomerular basement membrane), Goodpasture's syndrome, autoimmune hemolytic anemia (antibodies to erythrocytes), Hashimoto's disease (antibodies to thyroid), pernicious anemia (antibodies to intrinsic factor), idiopathic thrombocytopenic purpura (antibodies to platelets), Grave's disease, and Addison's disease (antibodies to thyroglobulin), and the like.

The autoantigens associated with a number of these diseases have been identified. For example, in experimentally induced autoimmune diseases, antigens involved in pathogenesis have been characterized: in arthritis in rat and mouse,

native type-II collagen is identified in collagen-induced arthritis, and mycobacterial heat shock protein in adjuvant arthritis; thyroglobulin has been identified in experimental allergic thyroiditis (EAT) in mouse; acetyl choline receptor (AChR) in experimental allergic myasthenia gravis (EAMG); and myelin basic protein (MBP) and proteolipid protein (PLP) in experimental allergic encephalomyelitis (EAE) in mouse and rat. In addition, target antigens have been identified in humans: type-II collagen in human rheumatoid arthritis; and acetyl choline receptor in myasthenia gravis.

Peptides comprising the epitopes from these antigens are synthesized and then tested for their ability to bind to the appropriate MHC molecules in assays using, for example, purified class I molecules and radioiodonated peptides and/or cells expressing empty class I molecules by, for instance, immunofluorescent staining and flow microfluorometry, peptide-dependent class I assembly assays, and inhibition of CTL recognition by peptide competition. Those peptides that bind to the class I molecule are further evaluated for their ability to serve as targets for CTLs derived from infected or immunized individuals, as well as for their capacity to induce primary in vitro or in vivo CTL responses that can give rise to CTL populations capable of reacting with virally infected target cells or tumor cells as potential therapeutic agents.

The MHC class I antigens are encoded by the HLA-A, B, and C loci. HLA-A and B antigens are expressed at the cell surface at approximately equal densities, whereas the expression of HLA-C is significantly lower (perhaps as much as 10-fold lower). Each of these loci have a number of alleles. The peptide binding motifs of the invention are relatively specific for each allelic subtype.

For peptide-based vaccines, the peptides of the present invention preferably comprise a motif recognized by an MHC I molecule having a wide distribution in the human population. Since the MHC alleles occur at different frequencies within different ethnic groups and races, the choice of target MHC allele may depend upon the target population. Table 1 shows the frequency of various alleles at

the HLA-A locus products among different races. For instance, the majority of the Caucasoid population can be covered by peptides which bind to four HLA-A allele subtypes, specifically HLA-A2.1, A1, A3.2, and A24.1. Similarly, the majority of the Asian population is encompassed with the addition of peptides binding to a fifth allele HLA-A11.2.

TABLE 1

A Allele/Subtype	N(69)*	A(54)	C(502)
A1	10.1(7)	1.8(1)	27.4(138)
A2.1	11.5(8)	37.0(20)	39.8(199)
A2.2	10.1(7)	0	3.3(17)
A2.3	1.4(1)	5.5(3)	0.8(4)
A2.4	-	-	-
A2.5	-	-	-
A3.1	1.4(1)	0	0.2(0)
A3.2	5.7(4)	5.5(3)	21.5(108)
A11.1	0	5.5(3)	0
A11.2	5.7(4)	31.4(17)	8.7(44)
A11.3	0	3.7(2)	0
A23	4.3(3)	-	3.9(20)
A24	2.9(2)	27.7(15)	15.3(77)
A24.2	-	-	-
A24.3	-	-	-
A25	1.4(1)	-	6.9(35)
A26.1	4.3(3)	9.2(5)	5.9(30)
A26.2	7.2(5)	-	1.0(5)
A26V	-	3.7(2)	-
A28.1	10.1(7)	-	1.6(8)
A28.2	1.4(1)	-	7.5(38)
A29.1	1.4(1)	-	1.4(7)
A29.2	10.1(7)	1.8(1)	5.3(27)
A30.1	8.6(6)	-	4.9(25)
A30.2	1.4(1)	-	0.2(1)
A30.3	7.2(5)	-	3.9(20)
A31	4.3(3)	7.4(4)	6.9(35)
A32	2.8(2)	-	7.1(36)
Aw33.1	8.6(6)	-	2.5(13)
Aw33.2	2.8(2)	16.6(9)	1.2(6)
Aw34.1	1.4(1)	-	-
Aw34.2	14.5(10)	-	0.8(4)
Aw36	5.9(4)	-	-

Table compiled from B. DuPont, Immunobiology of HLA, Vol. I, Histocompatibility Testing 1987, Springer-Verlag, New York 1989.

* N - negroid; A = Asian; C = caucasoid. Numbers in parenthesis represent the number of individuals included in the analysis.

The nomenclature used to describe peptide compounds follows the conventional practice wherein the amino group is presented to the left (the N-terminus) and the carboxyl group to the right (the C-terminus) of each amino acid residue. In the formulae representing selected specific embodiments of the present invention, the amino- and carboxyl-terminal groups, although not specifically shown, are in the form they would assume at physiologic pH values, unless otherwise specified. In the amino acid structure formulae, each residue is generally represented by standard three letter or single letter designations. The L-form of an amino acid residue is represented by a capital single letter or a capital first letter of a three-letter symbol, and the D-form for those amino acids having D-forms is represented by a lower case single letter or a lower case three letter symbol. Glycine has no asymmetric carbon atom and is simply referred to as "Gly" or G.

The procedures used to identify peptides of the present invention generally follow the methods disclosed in Falk et al., Nature 351:290 (1991), which is incorporated herein by reference. Briefly, the methods involve large-scale isolation of MHC class I molecules, typically by immunoprecipitation or affinity chromatography, from the appropriate cell or cell line. Examples of other methods for isolation of the desired MHC molecule equally well known to the artisan include ion exchange chromatography, lectin chromatography, size exclusion, high performance ligand chromatography, and a combination of all of the above techniques.

In the typical case, immunoprecipitation is used to isolate the desired allele. A number of protocols can be used, depending upon the specificity of the antibodies used. For example, allele-specific mAb reagents can be used for the affinity purification of the HLA-A, HLA-B_1, and HLA-C molecules. Several mAb reagents for the isolation of HLA-A molecules are available. The monoclonal BB7.2 is suitable for isolating HLA-A2 molecules. Affinity columns prepared with

these mAbs using standard techniques are successfully used to purify the respective HLA-A allele products.

In addition to allele-specific mAbs, broadly reactive anti-HLA-A, B, C mAbs, such as W6/32 and B9.12.1, and one anti-HLA-B, C mAb, B1.23.2, could be used in alternative affinity purification protocols as described in the example section below.

The peptides bound to the peptide binding groove of the isolated MHC molecules are eluted typically using acid treatment. Peptides can also be dissociated from class I molecules by a variety of standard denaturing means, such as heat, pH, detergents, salts, chaotropic agents, or a combination thereof.

Peptide fractions are further separated from the MHC molecules by reversed-phase high performance liquid chromatography (HPLC) and sequenced. Peptides can be separated by a variety of other standard means well known to the artisan, including filtration, ultrafiltration, electrophoresis, size chromatography, precipitation with specific antibodies, ion exchange chromatography, isoelectrofocusing, and the like.

Sequencing of the isolated peptides can be performed according to standard techniques such as Edman degradation (Hunkapiller, M.W., et al., Methods Enzymol. 91, 399 [1983]). Other methods suitable for sequencing include mass spectrometry sequencing of individual peptides as previously described (Hunt, et al., Science 225:1261 (1992), which is incorporated herein by reference). Amino acid sequencing of bulk heterogenous peptides (e.g., pooled HPLC fractions) from different class I molecules typically reveals a characteristic sequence motif for each class I allele.

Definition of motifs specific for different class I alleles allows the identification of potential peptide epitopes from an antigenic protein whose amino acid sequence is known. Typically, identification of potential peptide epitopes is initially carried out using a computer to scan the amino acid sequence of a desired antigen for the presence of motifs. The epitopic sequences are then synthesized. The

capacity to bind MHC Class molecules is measured in a variety of different ways. One means is a Class I molecule binding assay as described in Example 4, below. Other alternatives described in the literature include inhibition of antigen presentation (Sette, et al., J. Immunol. 141:3893 (1991), in vitro assembly assays (Townsend, et al., Cell 62:285 (1990), and FACS based assays using mutated ells, such as RMA.S (Melief, et al., Eur. J. Immunol. 21:2963 (1991)).

Next, peptides that test positive in the MHC class I binding assay are assayed for the ability of the peptides to induce specific CTL responses in vitro. For instance, Antigen-presenting cells that have been incubated with a peptide can be assayed for the ability to induce CTL responses in responder cell populations. Antigen-presenting cells can be normal cells such as peripheral blood mononuclear cells or dendritic cells (Inaba, et al., J. Exp. Med. 166:182 (1987); Boog, Eur. J. Immunol. 18:219 [1988]).

Alternatively, mutant mammalian cell lines that are deficient in their ability to load class I molecules with internally processed peptides, such as the mouse cell lines RMA-S (Kärre, et al.. Nature, 319:675 (1986); Ljunggren, et al., Eur. J. Immunol. 21:2963-2970 (1991)), and the human somatic T cell hybrid, T-2 (Cerundolo, et al., Nature 345:449-452 (1990)) and which have been transfected with the appropriate human class I genes are conveniently used, when peptide is added to them, to test for the capacity of the peptide to induce in vitro primary CTL responses. Other eukaryotic cell lines which could be used include various insect cell lines such as mosquito larvae (ATCC cell lines CCL 125, 126, 1660, 1591, 6585, 6586), silkworm (ATTC CRL 8851), armyworm (ATCC CRL 1711), moth (ATCC CCL 80) and Drosophila cell lines such as a Schneider cell line (see Schneider J. Embryol. Exp. Morphol. 27:353-365 [1927]).

Peripheral blood lymphocytes are conveniently isolated following simple venipuncture or leukapheresis of normal donors or patients and used as the responder cell sources of CTL precursors. In one embodiment, the appropriate antigen-presenting cells are incubated with 10-100 μM of

peptide in serum-free media for 4 hours under appropriate culture conditions. The peptide-loaded antigen-presenting cells are then incubated with the responder cell populations in vitro for 7 to 10 days under optimized culture conditions. Positive CTL activation can be determined by assaying the cultures for the presence of CTLs that kill radiolabeled target cells, both specific peptide-pulsed targets as well as target cells expressing endogenously processed form of the relevant virus or tumor antigen from which the peptide sequence was derived.

Specificity and MHC restriction of the CTL is determined by testing against different peptide target cells expressing appropriate or inappropriate human MHC class I. The peptides that test positive in the MHC binding assays and give rise to specific CTL responses are referred to herein as immunogenic peptides.

The immunogenic peptides can be prepared synthetically, or by recombinant DNA technology or from natural sources such as whole viruses or tumors. Although the peptide will preferably be substantially free of other naturally occurring host cell proteins and fragments thereof, in some embodiments the peptides can be synthetically conjugated to native fragments or particles.

The polypeptides or peptides can be a variety of lengths, either in their neutral (uncharged) forms or in forms which are salts, and either free of modifications such as glycosylation, side chain oxidation, or phosphorylation or containing these modifications, subject to the condition that the modification not destroy the biological activity of the polypeptides as herein described.

Desirably, the peptide will be as small as possible while still maintaining substantially all of the biological activity of the large peptide. When possible, it may be desirable to optimize peptides of the invention to a length of about 8 to about 10 amino acid residues, commensurate in size with endogenously processed viral peptides or tumor cell peptides that are bound to MHC class I molecules on the cell surface.

Peptides having the desired activity may be modified as necessary to provide certain desired attributes, e.g., improved pharmacological characteristics, while increasing or at least retaining substantially all of the biological activity of the unmodified peptide to bind the desired MHC molecule and activate the appropriate T cell. For instance, the peptides may be subject to various changes, such as substitutions, either conservative or non-conservative, where such changes might provide for certain advantages in their use, such as improved MHC binding. By conservative substitutions is meant replacing an amino acid residue with another which is biologically and/or chemically similar, e.g., one hydrophobic residue for another, or one polar residue for another. The substitutions include combinations such as Gly, Ala; Val, Ile, Leu, Met; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. The effect of single amino acid substitutions may also be probed using D-amino acids. Such modifications may be made using well known peptide synthesis procedures, as described in e.g., Merrifield, Science 232:341-347 (1986), Barany and Merrifield, The Peptides, Gross and Meienhofer, eds. (N.Y., Academic Press), pp. 1-284 (1979); and Stewart and Young, Solid Phase Peptide Synthesis, (Rockford, Ill., Pierce), 2d Ed. (1984), incorporated by reference herein.

The peptides can also be modified by extending or decreasing the compound's amino acid sequence, e.g., by the addition or deletion of amino acids. The peptides or analogs of the invention can also be modified by altering the order or composition of certain residues, it being readily appreciated that certain amino acid residues essential for biological activity, e.g., those at critical contact sites or conserved residues, may generally not be altered without an adverse effect on biological activity. The non-critical amino acids need not be limited to those naturally occurring in proteins, such as L-α-amino acids, or their D-isomers, but may include non-natural amino acids as well, such as β-γ-δ-amino acids, as well as many derivatives of L-α-amino acids.

Typically, a series of peptides with single amino acid substitutions are employed to determine the effect of electrostatic charge, hydrophobicity, etc. on binding. For instance, a series of positively charged (e.g., Lys or Arg) or negatively charged (e.g., Glu) amino acid substitutions are made along the length of the peptide revealing different patterns of sensitivity towards various MHC molecules and T cell receptors. In addition, multiple substitutions using small, relatively neutral moieties such as Ala, Gly, Pro, or similar residues may be employed. The substitutions may be homo-oligomers or hetero-oligomers. The number and types of residues which are substituted or added depend on the spacing necessary between essential contact points and certain functional attributes which are sought (e.g., hydrophobicity versus hydrophilicity). Increased binding affinity for an MHC molecule or T cell receptor may also be achieved by such substitutions, compared to the affinity of the parent peptide. In any event, such substitutions should employ amino acid residues or other molecular fragments chosen to avoid, for example, steric and charge interference which might disrupt binding.

Amino acid substitutions are typically of single residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final peptide. Substitutional variants are those in which at least one residue of a peptide has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 2 when it is desired to finely modulate the characteristics of the peptide.

TABLE 2

Original Residue	Exemplary Substitution
Ala	Ser
Arg	Lys, His
Asn	Gln
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Lys; Arg
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; His
Met	Leu; Ile
Phe	Tyr; Trp
Ser	Thr
Thr	Ser
Trp	Tyr; Phe
Tyr	Trp; Phe
Val	Ile; Leu

Substantial changes in function (e.g., affinity for MHC molecules or T cell receptors) are made by selecting substitutions that are less conservative than those in Table 2, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the peptide backbone in the area of the substitution, for example as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in peptide properties will be those in which (a) hydrophilic residue, e.g. seryl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a residue having an electropositive side chain, e.g., lysl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (c) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

The peptides may also comprise isosteres of two or more residues in the immunogenic peptide. An isostere as defined here is a sequence of two or more residues that can be substituted for a second sequence because the steric conformation of the first sequence fits a binding site specific for the second sequence. The term specifically includes peptide backbone modifications well known to those skilled in the art. Such modifications include modifications of the amide nitrogen, the α-carbon, amide carbonyl, complete replacement of the amide bond, extensions, deletions or backbone crosslinks. See, generally, Spatola, Chemistry and Biochemistry of Amino Acids, peptides and Proteins, Vol. VII (Weinstein ed., 1983).

Modifications of peptides with various amino acid mimetics or unnatural amino acids are particularly useful in increasing the stability of the peptide in vivo. Stability can be assayed in a number of ways. For instance, peptidases and various biological media, such as human plasma and serum, have been used to test stability. See, e.g., Verhoef et al.,

Eur. J. Drug Metab. Pharmacokin. 11:291-302 (1986). Half life of the peptides of the present invention is conveniently determined using a 25% human serum (v/v) assay. The protocol is generally as follows. Pooled human serum (Type AB, non-heat inactivated) is delipidated by centrifugation before use. The serum is then diluted to 25% with RPMI tissue culture media and used to test peptide stability. At predetermined time intervals a small amount of reaction solution is removed and added to either 6% aqueous trichloracetic acid or ethanol. The cloudy reaction sample is cooled (4°C) for 15 minutes and then spun to pellet the precipitated serum proteins. The presence of the peptides is then determined by reversed-phase HPLC using stability-specific chromatography conditions.

The peptides of the present invention or analogs thereof which have CTL stimulating activity may be modified to provide desired attributes other than improved serum half life. For instance, the ability of the peptides to induce CTL activity can be enhanced by linkage to a sequence which contains at least one epitope that is capable of inducing a T helper cell response.

In some embodiments, the T helper peptide is one that is recognized by T helper cells in the majority of the population. This can be accomplished by selecting amino acid sequences that bind to many, most, or all of the MHC class II molecules. These are known as "loosely MHC-restricted" T helper sequences. Examples of amino acid sequences that are loosely MHC-restricted include sequences from antigens such as Tetanus toxin at positions 830-843 (QYIKANSKFIGITE), *Plasmodium falciparum* CS protein at positions 378-398 (DIEKKIAKMEKASSVFNVVNS), and Streptococcus 18kD protein at positions 1-16 (YGAVDSILGGVATYGAA).

Alternatively, it is possible to prepare synthetic peptides capable of stimulating T helper lymphocytes, in a loosely MHC-restricted fashion, using amino acid sequences not found in nature. These synthetic compounds called Pan-DR-binding epitope (PADRE) are designed on the basis of

their binding activity to most, HLA-DR (human MHC class II) molecules (*see*, copending application USSN 08/121,101).

Particularly preferred immunogenic peptides/T helper conjugates are linked by a spacer molecule. The spacer is typically comprised of relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological conditions. The spacers are typically selected from, e.g., Ala, Gly, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present spacer need not comprise the same residues and thus may be a hetero- or homo-oligomer. When present, the spacer will usually be at least one or two residues, more usually three to six residues. Alternatively, the CTL peptide may be linked to the T helper peptide without a spacer.

The immunogenic peptide may be linked to the T helper peptide either directly or via a spacer either at the amino or carboxy terminus of the CTL peptide. The amino terminus of either the immunogenic peptide or the T helper peptide may be acylated. Exemplary T helper peptides include tetanus toxoid 830-843, influenza 307-319, malaria circumsporozoite 382-398 and 378-389.

In some embodiments it may be desirable to include in the pharmaceutical compositions of the invention at least one component which primes CTL. Lipids have been identified as agents capable of priming CTL <u>in vivo</u> against viral antigens. For example, palmitic acid residues can be attached to the alpha and epsilon amino groups of a Lys residue and then linked, e.g., via one or more linking residues such as Gly, Gly-Gly-, Ser, Ser-Ser, or the like, to an immunogenic peptide. The lipidated peptide can then be injected directly in a micellar form, incorporated into a liposome or emulsified in an adjuvant, e.g., incomplete Freund's adjuvant. In a preferred embodiment a particularly effective immunogen comprises palmitic acid attached to alpha and epsilon amino groups of Lys, which is attached via linkage, e.g., Ser-Ser, to the amino terminus of the immunogenic peptide.

As another example of lipid priming of CTL responses, E. coli lipoproteins, such as tripalmitoyl-S-glycerylcysteinlyseryl-serine (P_3CSS) can be used to prime virus specific CTL when covalently attached to an appropriate peptide. See, Deres et al., Nature 342:561-564 (1989), incorporated herein by reference. Peptides of the invention can be coupled to P_3CSS, for example, and the lipopeptide administered to an individual to specifically prime a CTL response to the target antigen. Further, as the induction of neutralizing antibodies can also be primed with P_3CSS conjugated to a peptide which displays an appropriate epitope, the two compositions can be combined to more effectively elicit both humoral and cell-mediated responses to infection.

In addition, additional amino acids can be added to the termini of a peptide to provide for ease of linking peptides one to another, for coupling to a carrier support, or larger peptide, for modifying the physical or chemical properties of the peptide or oligopeptide, or the like. Amino acids such as tyrosine, cysteine, lysine, glutamic or aspartic acid, or the like, can be introduced at the C- or N-terminus of the peptide or oligopeptide. Modification at the C terminus in some cases may alter binding characteristics of the peptide. In addition, the peptide or oligopeptide sequences can differ from the natural sequence by being modified by terminal-NH_2 acylation, e.g., by alkanoyl (C_1-C_{20}) or thioglycolyl acetylation, terminal-carboxyl amidation, e.g., ammonia, methylamine, etc. In some instances these modifications may provide sites for linking to a support or other molecule.

The peptides of the invention can be prepared in a wide variety of ways. Because of their relatively short size, the peptides can be synthesized in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. See, for example, Stewart and Young, Solid Phase Peptide Synthesis, 2d. ed., Pierce Chemical Co. (1984), supra.

Alternatively, recombinant DNA technology may be employed wherein a nucleotide sequence which encodes an immunogenic peptide of interest is inserted into an expression vector, transformed or transfected into an appropriate host cell and cultivated under conditions suitable for expression. These procedures are generally known in the art, as described generally in Sambrook et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, New York (1982), which is incorporated herein by reference. Thus, fusion proteins which comprise one or more peptide sequences of the invention can be used to present the appropriate T cell epitope.

As the coding sequence for peptides of the length contemplated herein can be synthesized by chemical techniques, for example, the phosphotriester method of Matteucci et al., J. Am. Chem. Soc. 103:3185 (1981), modification can be made simply by substituting the appropriate base(s) for those encoding the native peptide sequence. The coding sequence can then be provided with appropriate linkers and ligated into expression vectors commonly available in the art, and the vectors used to transform suitable hosts to produce the desired fusion protein. A number of such vectors and suitable host systems are now available. For expression of the fusion proteins, the coding sequence will be provided with operably linked start and stop codons, promoter and terminator regions and usually a replication system to provide an expression vector for expression in the desired cellular host. For example, promoter sequences compatible with bacterial hosts are provided in plasmids containing convenient restriction sites for insertion of the desired coding sequence. The resulting expression vectors are transformed into suitable bacterial hosts. Of course, yeast or mammalian cell hosts may also be used, employing suitable vectors and control sequences.

The peptides of the present invention and pharmaceutical and vaccine compositions thereof are useful for administration to mammals, particularly humans, to treat and/or prevent viral infection and cancer. Examples of

diseases which can be treated using the immunogenic peptides of the invention include prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma, cervical carcinoma, lymphoma, CMV and condlyloma acuminatum.

For pharmaceutical compositions, the immunogenic peptides of the invention are administered to an individual already suffering from cancer or infected with the virus of interest. Those in the incubation phase or the acute phase of infection can be treated with the immunogenic peptides separately or in conjunction with other treatments, as appropriate. In therapeutic applications, compositions are administered to a patient in an amount sufficient to elicit an effective CTL response to the virus or tumor antigen and to cure or at least partially arrest symptoms and/or complications. An amount adequate to accomplish this is defined as "therapeutically effective dose." Amounts effective for this use will depend on, e.g., the peptide composition, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician, but generally range for the initial immunization (that is for therapeutic or prophylactic administration) from about 1.0 μg to about 5000 μg of peptide for a 70 kg patient, followed by boosting dosages of from about 1.0 μg to about 1000 μg of peptide pursuant to a boosting regimen over weeks to months depending upon the patient's response and condition by measuring specific CTL activity in the patient's blood. It must be kept in mind that the peptides and compositions of the present invention may generally be employed in serious disease states, that is, life-threatening or potentially life threatening situations. In such cases, in view of the minimization of extraneous substances and the relative nontoxic nature of the peptides, it is possible and may be felt desirable by the treating physician to administer substantial excesses of these peptide compositions.

For therapeutic use, administration should begin at the first sign of viral infection or the detection or surgical

removal of tumors or shortly after diagnosis in the case of acute infection. This is followed by boosting doses until at least symptoms are substantially abated and for a period thereafter. In chronic infection, loading doses followed by boosting doses may be required.

Treatment of an infected individual with the compositions of the invention may hasten resolution of the infection in acutely infected individuals. For those individuals susceptible (or predisposed) to developing chronic infection the compositions are particularly useful in methods for preventing the evolution from acute to chronic infection. Where the susceptible individuals are identified prior to or during infection, for instance, as described herein, the composition can be targeted to them, minimizing need for administration to a larger population.

The peptide compositions can also be used for the treatment of chronic infection and to stimulate the immune system to eliminate virus-infected cells in carriers. It is important to provide an amount of immuno-potentiating peptide in a formulation and mode of administration sufficient to effectively stimulate a cytotoxic T cell response. Thus, for treatment of chronic infection, a representative dose is in the range of about 1.0 μg to about 5000 μg, preferably about 5 μg to 1000 μg for a 70 kg patient per dose. Immunizing doses followed by boosting doses at established intervals, e.g., from one to four weeks, may be required, possibly for a prolonged period of time to effectively immunize an individual. In the case of chronic infection, administration should continue until at least clinical symptoms or laboratory tests indicate that the viral infection has been eliminated or substantially abated and for a period thereafter.

The pharmaceutical compositions for therapeutic treatment are intended for parenteral, topical, oral or local administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration which comprise a solution of the immunogenic peptides

dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.8% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

The concentration of CTL stimulatory peptides of the invention in the pharmaceutical formulations can vary widely, i.e., from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected.

The peptides of the invention may also be administered via liposomes, which serve to target the peptides to a particular tissue, such as lymphoid tissue, or targeted selectively to infected cells, as well as increase the half-life of the peptide composition. Liposomes include emulsions, foams, micelles, insoluble monolayers, liquid crystals, phospholipid dispersions, lamellar layers and the like. In these preparations the peptide to be delivered is incorporated as part of a liposome, alone or in conjunction with a molecule which binds to, e.g., a receptor prevalent among lymphoid cells, such as monoclonal antibodies which bind to the CD45 antigen, or with other therapeutic or immunogenic compositions. Thus, liposomes either filled or decorated with a desired peptide of the invention can be directed to the site of lymphoid cells, where the liposomes then deliver the selected therapeutic/immunogenic peptide compositions.

Liposomes for use in the invention are formed from standard vesicle-forming lipids, which generally include neutral and negatively charged phospholipids and a sterol, such as cholesterol. The selection of lipids is generally guided by consideration of, e.g., liposome size, acid lability and stability of the liposomes in the blood stream. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka et al., Ann. Rev. Biophys. Bioeng. 9:467 (1980), U.S. Patent Nos. 4,235,871, 4,501,728, 4,837,028, and 5,019,369, incorporated herein by reference.

For targeting to the immune cells, a ligand to be incorporated into the liposome can include, e.g., antibodies or fragments thereof specific for cell surface determinants of the desired immune system cells. A liposome suspension containing a peptide may be administered intravenously, locally, topically, etc. in a dose which varies according to, inter alia, the manner of administration, the peptide being delivered, and the stage of the disease being treated.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient, that is, one or more peptides of the invention, and more preferably at a concentration of 25%-75%.

For aerosol administration, the immunogenic peptides are preferably supplied in finely divided form along with a surfactant and propellant. Typical percentages of peptides are 0.01%-20% by weight, preferably 1%-10%. The surfactant must, of course, be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an

aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. The surfactant may constitute 0.1%-20% by weight of the composition, preferably 0.25-5%. The balance of the composition is ordinarily propellant. A carrier can also be included, as desired, as with, e.g., lecithin for intranasal delivery.

In another aspect the present invention is directed to vaccines which contain as an active ingredient an immunogenically effective amount of an immunogenic peptide as described herein. The peptide(s) may be introduced into a host, including humans, linked to its own carrier or as a homopolymer or heteropolymer of active peptide units. Such a polymer has the advantage of increased immunological reaction and, where different peptides are used to make up the polymer, the additional ability to induce antibodies and/or CTLs that react with different antigenic determinants of the virus or tumor cells. Useful carriers are well known in the art, and include, e.g., thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino acids such as poly(lysine:glutamic acid), influenza, hepatitis B virus core protein, hepatitis B virus recombinant vaccine and the like. The vaccines can also contain a physiologically tolerable (acceptable) diluent such as water, phosphate buffered saline, or saline, and further typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art. And, as mentioned above, CTL responses can be primed by conjugating peptides of the invention to lipids, such as P_3CSS. Upon immunization with a peptide composition as described herein, via injection, aerosol, oral, transdermal or other route, the immune system of the host responds to the vaccine by producing large amounts of CTLs specific for the desired antigen, and the host becomes at least partially immune to later infection, or resistant to developing chronic infection.

Vaccine compositions containing the peptides of the invention are administered to a patient susceptible to or

otherwise at risk of viral infection or cancer to elicit an immune response against the antigen and thus enhance the patient's own immune response capabilities. Such an amount is defined to be an "immunogenically effective dose." In this use, the precise amounts again depend on the patient's state of health and weight, the mode of administration, the nature of the formulation, etc., but generally range from about 1.0 μg to about 5000 μg per 70 kilogram patient, more commonly from about 10 μg to about 500 μg mg per 70 kg of body weight.

In some instances it may be desirable to combine the peptide vaccines of the invention with vaccines which induce neutralizing antibody responses to the virus of interest, particularly to viral envelope antigens.

For therapeutic or immunization purposes, the peptides of the invention can also be expressed by attenuated viral hosts, such as vaccinia or fowlpox. This approach involves the use of vaccinia virus as a vector to express nucleotide sequences that encode the peptides of the invention. Upon introduction into an acutely or chronically infected host or into a non-infected host, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits a host CTL response. Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Patent No. 4,722,848, incorporated herein by reference. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (Nature 351:456-460 (1991)) which is incorporated herein by reference. A wide variety of other vectors useful for therapeutic administration or immunization of the peptides of the invention, e.g., Salmonella typhi vectors and the like, will be apparent to those skilled in the art from the description herein.

Antigenic peptides may be used to elicit CTL ex vivo, as well. The resulting CTL, can be used to treat chronic infections (viral or bacterial) or tumors in patients that do not respond to other conventional forms of therapy, or will not respond to a peptide vaccine approach of therapy. Ex vivo CTL responses to a particular pathogen (infectious agent or tumor antigen) are induced by incubating in tissue culture the

patient's CTL precursor cells (CTLp) together with a source of antigen-presenting cells (APC) and the appropriate immunogenic peptide. After an appropriate incubation time (typically 1-4 weeks), in which the CTLp are activated and mature and expand into effector CTL, the cells are infused back into the patient, where they will destroy their specific target cell (an infected cell or a tumor cell).

The peptides may also find use as diagnostic reagents. For example, a peptide of the invention may be used to determine the susceptibility of a particular individual to a treatment regimen which employs the peptide or related peptides, and thus may be helpful in modifying an existing treatment protocol or in determining a prognosis for an affected individual. In addition, the peptides may also be used to predict which individuals will be at substantial risk for developing chronic infection.

The following examples are offered by way of illustration, not by way of limitation.

Example 1

Class I antigen isolation

A flow diagram of an HLA-A antigen purification scheme is presented in Figure 1. Briefly, the cells bearing the appropriate allele were grown in large batches (6-8 liters yielding ~5 x 10^9 cells), harvested by centrifugation and washed. All cell lines were maintained in RPMI 1640 media (Sigma) supplemented with 10% fetal bovine serum (FBS) and antibiotics. For large-scale cultures, cells were grown in roller bottle culture in RPMI 1640 with 10% FBS or with 10% horse serum and antibiotics. Cells were harvested by centrifugation at 1500 RPM IEC-CRU5000 centrifuge with 259 rotor and washed three times with phosphate-buffered saline (PBS)(0.01 M PO_4, 0.154 M NaCl, pH 7.2).

Cells were pelleted and stored at -70°C or treated with detergent lysing solution to prepare detergent lysates. Cell lysates were prepared by the addition of stock detergent solution [1% NP-40 (Sigma) or Renex 30 (Accurate Chem. Sci. Corp., Westbury, NY 11590), 150 mM NaCl, 50 mM Tris, pH 8.0] to the cell pellets (previously counted) at a ratio of 50-100 x 10^6 cells per ml detergent solution. A cocktail of protease inhibitors was added to the premeasured volume of stock detergent solution immediately prior to the addition to the cell pellet. Addition of the protease inhibitor cocktail produced final concentrations of the following: phenylmethylsulfonyl fluoride (PMSF), 2 mM; aprotinin, 5 μg/ml; leupeptin, 10 μg/ml; pepstatin, 10 μg/ml; iodoacetamide, 100 μM; and EDTA, 3 ng/ml. Cell lysis was allowed to proceed at 4°C for 1 hour with periodic mixing. Routinely 5-10 x 10^9 cells were lysed in 50-100 ml of detergent solution. The lysate was clarified by centrifugation at 15,000 x g for 30 minutes at 4°C and subsequent passage of the supernatant fraction through a 0.2 μ filter unit (Nalgene).

The HLA-A antigen purification was achieved using affinity columns prepared with mAb-conjugated Sepharose beads. For antibody production, cells were grown in RPMI with 10% FBS in large tissue culture flasks (Corning 25160-225).

Antibodies were purified from clarified tissue culture medium by ammonium sulfate fractionation followed by affinity chromatography on protein-A-Sepharose (Sigma). Briefly, saturated ammonium sulfate was added slowly with stirring to the tissue culture supernatant to 45% (volume to volume) overnight at 4°C to precipitate the immunoglobulins. The precipitated proteins were harvested by centrifugation at 10,000 x g for 30 minutes. The precipitate was then dissolved in a minimum volume of PBS and transferred to dialysis tubing (Spectro/Por 2, Mol. wt. cutoff 12,000-14,000, Spectum Medical Ind.). Dialysis was against PBS (≥20 times the protein solution volume) with 4-6 changes of dialysis buffer over a 24-48 hour period at 4°C. The dialyzed protein solution was clarified by centrifugation (10,000 x g for 30 minutes) and the pH of the solution adjusted to pH 8.0 with 1N NaOH. Protein-A-Sepharose (Sigma) was hydrated according to the manufacturer's instructions, and a protein-A-Sepharose column was prepared. A column of 10 ml bed volume typically binds 50-100 mg of mouse IgG.

The protein sample was loaded onto the protein-A-Sepharose column using a peristaltic pump for large loading volumes or by gravity for smaller volumes (<100 ml). The column was washed with several volumes of PBS, and the eluate was monitored at A280 in a spectrophotometer until base line was reached. The bound antibody was eluted using 0.1 M citric acid at suitable pH (adjusted to the appropriate pH with 1N NaOH). For mouse IgG-1 pH 6.5 was used for IgG2a pH 4.5 was used and for IgG2b and IgG3 pH 3.0 was used. 2 M Tris base was used to neutralize the eluate. Fractions containing the antibody (monitored by A280) were pooled, dialyzed against PBS and further concentrated using an Amicon Stirred Cell system (Amicon Model 8050 with YM30 membrane). The anti-A2 mAb, BB7.2, was useful for affinity purification.

The HLA-A antigen was purified using affinity columns prepared with mAb-conjugated Sepharose beads. The affinity columns were prepared by incubating protein-A-Sepharose beads (Sigma) with affinity-purified mAb as described above. Five to 10 mg of mAb per ml of bead is the preferred ratio. The

mAb bound beads were washed with borate buffer (borate buffer: 100 mM sodium tetraborate, 154 mM NaCl, pH 8.2) until the washes show A280 at based line. Dimethyl pimelimidate (20 mM) in 200 mM triethanolamine was added to covalently crosslink the bound mAb to the protein-A-Sepharose (Schneider et al., J. Biol. Chem. 257:10766 (1982). After incubation for 45 minutes at room temperature on a rotator, the excess crosslinking reagent was removed by washing the beads twice with 10-20 ml of 20 mM ethanolamine, pH 8.2. Between each one the slurry was placed on a rotator for 5 minutes at room temperature. The beads were washed with borate buffer and with PBS plus 0.02% sodium azide.

The cell lysate (5-10 x 10^9 cell equivalents) was then slowly passed over a 5-10 ml affinity column (flow rate of 0.1-0.25 ml per minute) to allow the binding of the antigen to the immobilized antibody. After the lysate was allowed to pass through the column, the column was washed sequentially with 20 column volumes of detergent stock solution plus 0.1% sodium dodecyl sulfate, 20 column volumes of 0.5 M NaCl, 20 mM Tris, pH 8.0, and 10 column volumes of 20 mM Tris, pH 8.0. The HLA-A antigen bound to the mAb was eluated with a basic buffer solution (50 mM diethylamine in water). As an alternative, acid solutions such as 0.15-0.25 M acetic acid were also used to elute the bound antigen. An aliquot of the eluate (1/50) was removed for protein quantification using either a colorimetric assay (BCA assay, Pierce) or by SDS-PAGE, or both. SDS-PAGE analysis was performed as described by Laemmli (Laemmli, U.K., Nature 227:680 (1970)) using known amounts of bovine serum albumin (Sigma) as a protein standard. Allele specific antibodies were used to purify the specific MHC molecule. In the case of HLA-A2, the mAb BB7.2 was used.

Example 2

Isolation and sequencing of naturally processed peptides

For the HLA-A preparations derived from the base (50 mM diethylamine) elution protocol, the eluate was immediately neutralized with 1 N acetic acid to pH 7.0-7.5. The neutralized eluate was concentrated to a volume of 1-2 ml in

an Amicon stirred cell [Model 8050, with YM3 membranes (Amicon)]. Ten ml of ammonium acetate (0.01 M, pH 8.0) was added to the concentrator to remove the non-volatile salts, and the sample was concentrated to approximately 1 ml. A small sample (1/50) was removed for protein quantitation as described above. The remainder was recovered into a 15 ml polypropylene conical centrifuge tube (Falcon, 2097) (Becton Dickinson). Glacial acetic acid was added to obtain a final concentration of 10% acetic acid. The acidified sample was placed in a boiling water bath for 5 minutes to allow for the dissociation of the bound peptides. The sample was cooled on ice, returned to the concentrator and the filtrate was collected. Additional aliquots of 10% acetic acid (1-2 ml) were added to the concentrator, and this filtrate was pooled with the original filtrate. Finally, 1-2 ml of distilled water was added to the concentrator, and this filtrate was pooled as well.

The retentate contains the bulk of the HLA-A heavy chain and β_2-microglobulin, while the filtrate contains the naturally processed bound peptides and other components with molecular weights less than about 3000. The pooled filtrate material was lyophilized in order to concentrate the peptide fraction. The sample was then ready for further analysis.

For HPLC (high performance liquid chromatography) separation of the peptide fractions, the lyophilized sample was dissolved in 50 μl of distilled water, or into 0.1% trifluoracetic acid (TFA) (Applied Biosystems) in water and injected to a C18 reverse-phase narrow bore column (Beckman C18 Ultrasphere, 10 x 250 mm), using a gradient system described by Stone and Williams (Stone, K.L. and Williams K.R., in, Macromolecular Sequencing and Synthesis; Selected Methods and Applications, A.R. Liss, New York, 1988, pp. 7-24. Buffer A was 0.06% TFA in water (Burdick-Jackson) and buffer B was 0.052% TFA in 80% acetonitrile (Burdick-Jackson). The flow rate was 0.250 ml/minute with the following gradient: 0-60 min., 2-37.5% B; 60-95 min., 37.5-75% B; 95-105 min., 75-98% B. The Gilson narrow bore HPLC configuration is

particularly useful for this purpose, although other configurations work equally well.

A large number of peaks were detected by absorbance at 214 nm, many of which appear to be of low abundance. Whether a given peak represents a single peptide or a peptide mixture was not determined. Pooled fractions were then sequenced to determine motifs specific for each allele as described below.

Pooled peptide fractions, prepared as described above were analyzed by automated Edman sequencing using the Applied Biosystems Model 477A automated sequencer. The sequencing method is based on the technique developed by Pehr Edman in the 1950s for the sequential degradation of proteins and peptides to determine the sequence of the constituent amino acids.

The protein or peptide to be sequenced was held by a 12-mm diameter porous glass fiber filter disk in a heated, argon-purged reaction chamber. The filter was generally pre-treated with BioBrene Plus™ and then cycled through one or more repetitions of the Edman reaction to reduce contaminants and improve the efficiency of subsequent sample sequencing. Following the pre-treatment of the filter, a solution of the sample protein or peptide (10 pmol-5 nmol range) was loaded onto the glass filter and dried. Thus, the sample was left embedded in the film of the pre-treated disk. Covalent attachment of the sample to the filter was usually not necessary because the Edman chemistry utilized relatively apolar solvents, in which proteins and peptides are poorly soluble.

Briefly, the Edman degradation reaction has three steps: coupling, cleavage, and conversion. In coupling step, phenylisothiocyanate (PITC) is added. The PITC reacts quantitatively with the free amino-terminal amino acid of the protein to form the phenylthiocarbamyl-protein in a basic environment. After a period of time for the coupling step, the excess chemicals are extracted and the highly volatile organic acid, trifluoroacetic acid, TFA, is used to cleave the PITC-coupled amino acid residue from the amino terminus of the protein yielding the anilinothiazolinone (ATZ) derivative of

the amino acid. The remaining protein/peptide is left with a new amino terminus and is ready for the next Edman cycle. The ATZ amino acid is extracted and transferred to a conversion flask, where upon addition of 25% TFA in water, the ATZ amino acid is converted to the more stable phenylthiohydantoin (PTH) amino acid that can be identified and quantified following automatic injection into the Model 120 PTH Analyzer which uses a microbore C-18 reverse-phase HPLC column for the analysis.

In the present procedures, peptide mixtures were loaded onto the glass filters. Thus, a single amino acid sequence usually does not result. Rather, mixtures of amino acids in different yield are found. When the particular residue is conserved among the peptides being sequenced, increased yield for that amino acid is observed.

Example 3

Definition of an A2.1 specific motif

In one case, pooled peptide fractions prepared as described in Example 2 above were obtained from HLA-A2.1 homozygous cell lines, for example, JY. The pooled fractions were HPLC fractions corresponding to 7% to 45% CH_3CN. For this class I molecule, this region of the chromatogram was most abundant in peptides. Data from independent experiments were averaged as described below.

The amino acid sequence analyses from four independent experiments were analyzed and the results are shown in Table 3. For each position except the first, the data were analyzed by modifying the method described by Falk et al., supra, to allow for comparison of experiments from different HLA types. This modified procedure yielded quantitative yet standardized values while allowing the averaging of data from different experiments involving the same HLA type.

The raw sequenator data was converted to a simple matrix of 10 rows (each representing one Edman degradation cycle) and 16 columns (each representing one of the twenty amino acids; W, C, R and H were eliminated for technical reasons. The data corresponding to the first row (first cycle) was not considered further because, this cycle is

usually heavily contaminated by free amino acids.). The values of each row were summed to yield a total pmoles value for that particular cycle. For each row, values for each amino acid were then divided by the corresponding total yield value, to determine what fraction of the total signal is attributable to each amino acid at each cycle. By doing so, an "Absolute Frequency" table was generated. This absolute frequency table allows correction for the declining yields of each cycle.

TABLE 3

A2.1:POOL SEQUENCING FREQUENCY

	pos. 1	pos. 2	pos. 3	pos. 4	pos. 5	pos. 6	pos. 7	pos. 8	pos. 9	pos. 10
A	-	0.65	1.25	0.85	0.95	0.77	1.21	1.16	1.15	1.25
G	-	0.84	0.96	1.29	1.22	0.89	0.78	1.05	0.98	1.48
D	-	0.84	1.11	1.70	1.03	0.83	0.82	0.84	0.82	1.19
E	-	0.38	0.59	1.73	1.10	0.82	1.05	1.45	0.87	0.88
R	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	-	-	-
K	-	0.63	0.65	0.89	1.66	1.09	0.89	1.35	0.82	0.87
L	-	2.66*	1.11	0.45	0.57	1.00	0.69	0.59	0.92	0.77
V	-	0.78	0.69	0.60	0.79	1.38	1.24	0.84	1.69	1.27
I	-	1.06	1.20	0.53	0.93	1.49	1.15	0.76	0.88	0.54
M	-	1.93	1.91	0.62	0.71	0.68	0.88	0.54	0.73	0.22
Y	-	0.28	1.41	0.65	1.32	0.78	1.34	1.21	1.00	0.79
F	-	0.76	1.46	0.69	1.16	1.00	1.07	1.09	0.78	0.73
W	-	-	-	-	-	-	-	-	-	-
Q	-	0.60	0.84	0.92	0.95	0.90	1.16	1.63	1.00	1.00
N	-	0.39	0.76	1.17	1.28	1.08	1.07	1.28	0.96	0.42
S	-	1.13	1.50	1.33	0.87	0.77	0.71	0.92	0.77	0.58
T	-	0.62	0.90	0.94	0.95	1.21	1.07	1.60	0.71	0.57
C	-	-	-	-	-	-	-	-	-	-
P	-	0.54	0.78	1.44	1.15	1.09	1.30	0.87	0.81	1.01

* [illegible]

Starting from the absolute frequency table, a "relative frequency" table was then generated to allow comparisons among different amino acids. To do so the data from each column was summed, and then averaged. Then, each value was divided next by the average column value to obtain relative frequency values. These values quantitate, in a standardized manner, increases and decreases per cycle, for each of the different sixteen amino acid types. Tables generated from data from different experiments can thus be added together to generate average relative frequency values (and their standard deviations). All standard deviations can then be averaged, to estimate a standard deviation value applicable to the samples from each table. Any particular value exceeding 1.00 by more than two standard deviations is considered to correspond to a significant increase.

Example 4

Quantitative Binding Assays

Using isolated MHC molecules prepared as described in Example 2, above, quantitative binding assays were performed. Briefly, indicated amounts of MHC as isolated above were incubated in 0.05% NP40-PBS with ~5 nM of radiolabeled peptides in the presence of 1-3 μM β_2M and a cocktail of protease inhibitors (final concentrations 1 mM PMSF, 1.3 mM 1.10 Phenanthroline, 73 μM Pepstatin A, 8 mM EDTA, 200 μM N-α-p-tosyl-L-Lysine Chloromethyl ketone). After various times, free and bound peptides were separated by TSK 2000 gel filtration, as described previously in A. Sette et al., J. Immunol. 148:844 (1992), which is incorporated herein by reference. Peptides were labeled by the use of the Chloramine T method Buus et al., Science 235:1352 (1987), which is incorporated herein by reference.

The HBc 18-27 peptide HLA binding peptide was radiolabeled and offered (5-10 nM) to 1 μM purified HLA A2.1. After two days at 23°C in presence of a cocktail of protease inhibitors and 1-3 μM purified human β_2M, the percent of MHC class I bound radioactivity was measured by size exclusion chromatography, as previously described for class II peptide

binding assays in Sette et al., in Seminars in Immunology, Vol. 3, Gefter, ed. (W.B. Saunders, Philadelphia, 1991), pp 195-202, which is incorporated herein by reference. Using this protocol, high binding (95%) was detected in all cases in the presence of purified HLA A2.1 molecules.

To explore the specificity of binding, we determined whether the binding was inhibitable by excess unlabeled peptide, and if so, what the 50% inhibitory concentration (IC50%) might be. The rationale for this experiment was threefold. First, such an experiment is crucial in order to demonstrate specificity. Second, a sensitive inhibition assay is the most viable alternative for a high throughput quantitative binding assay. Third, inhibition data subjected to Scatchard analysis can give quantitative estimates of the equilibrium constant (K) of interaction and the fraction of receptor molecules capable of binding ligand (% occupancy). For instance, in analysis of an inhibition curve for the interaction of the peptide HBc 18-27 with A2.1, the IC50% was determined to be 25 nM. Further experiments were conducted to obtain Scatchard plots. For HBc 18-27/A2.1, six different experiments using six independent MHC preparations yielded a K_D of 15.5 ± 9.9 x 10^{-9} M and occupancy values of 6.2% (±1.4).

Several reports have demonstrated that class I molecules, unlike class II, are highly selective with regard to the size of the peptide epitope that they recognize. The optimal size varies between 8 and 10 residues for different peptides and different class I molecules, although MHC binding peptides as long as 13 residues have been identified. To verify the stringent size requirement, a series of N- and C-terminal truncation/extension analogs of the peptide HBc 18-27 were synthesized and tested for A2.1 binding. Previous studies had demonstrated that the optimal size for CTL recognition of this peptide was the 10-mer HBc18-27 (Sette et al. supra). It was found that removal or addition of a residue at the C terminus of the molecule resulted in a 30 to 100-fold decrease in binding capacity. Further removal or addition of another residue completely obliterated binding. Similarly, at the N-terminus of the molecule, removal or

deletion of one residue from the optimal HBc 18-27 peptide completely abrogated A2.1 binding.

Throughout this disclosure, results have been expressed in terms of IC50's. Given the conditions in which our assays are run (i.e., limiting MHC and labeled peptide concentrations), these values approximate K_D values. It should be noted that IC50 values can change, often dramatically, if the assay conditions are varied, and depending on the particular reagents used (e.g., Class I preparation, etc.). For example, excessive concentrations of MHC will increase the apparent measured IC50 of a given ligand.

An alternative way of expressing the binding data, to avoid these uncertainties, is as a relative value to a reference peptide. The reference peptide is included in every assay. As a particular assay becomes more, or less, sensitive, the IC50's of the peptides tested may change somewhat. However, the binding relative to the reference peptide will not change. For example, in an assay run under conditions such that the IC50 of the reference peptide increases 10-fold, all IC50 values will also shift approximately ten-fold. Therefore, to avoid ambiguities, the assessment of whether a peptide is a good, intermediate, weak, or negative binder should be based on it's IC50, relative to the IC50 of the standard peptide.

The reference peptide for the HLA-A2.1 assays described herein is referred to as 941.01 having a sequence of FLPSDYFPSV. An average IC50 of 5 (nM) was observed under the assay conditions utilized.

If the IC50 of the standard peptide measured in a particular assay is different from that reported in the table, then it should be understood that the threshold values used to determine good, intermediate, weak, and negative binders should be modified by a corresponding factor. For example, if in an A2.1 binding assay, the IC50 of the A2.1 standard (941.01) were to be measured as 8 nM instead of 5 nM, then a peptide ligand would be called a good binder only if it had an

IC50 of less than 80 nM (i.e., 8nM x 0.1), instead of the usual cut-off value of 50 nM.

Example 5

HLA-A2.1 Binding Motif and Algorithm

The structural requirements for peptide binding to A2.1 have been defined for both, 9-mer and 10-mer peptides. Two approaches have been used. The first approach referred to as the "poly-A approach" uses a panel of single amino acid substitutions of a 9-mer prototype poly-A binder (ALAKAAAAV) that is tested for A2.1 binding using the methods of Example 4 above to examine the degree of degeneracy of the anchor-positions and the possible influence of non-anchor positions on A2.1 binding.

The second approach, the "Motif-Library approach", uses a large library of peptides selected from sequences of potential target molecules of viral and tumor origin and tested for A2.1 binding using the methods in Example 4 above. The frequencies by which different amino-acids occured at each position in good binders and non-binders were analysed to further define the role of non-anchor positions in 9-mers and 10-mers.

A2.1 binding of peptide 9-mers

Poly A Approach A poly-A 9-mer peptide, containing the A2.1 motif L (Leu) in position 2 and V (Val) in position 9 was chosen as a prototype binder. A K (Lys) residue was included in position 4 to increase solubility. A panel of 91 single amino-acid substitution analogues of the prototype parental 9-mer was synthesized and tested for A2.1 binding (Table 4). Shaded areas mark analogs with a greater than 10-fold reduction in binding capacity relative to the parental peptide. A reduction in binding greater than 100-fold is indicated by hyphenation.

Anchor-Positions 2 and 9 in poly-A Analogs The effect of single-amino-acid substitutions at the anchor positions 2 and 9 was examined first. Most substitutions in these positions had profound detrimental effects on binding

capacity, thus confirming their role for binding. More specifically, in position 2 only L and M bound within a 10-fold range ("preferred residues"). Residues with similar characteristics, such as I, V, A, and T were tolerated, but bound 10 to 100-fold less strongly than the parental peptide. All the remaining substitutions (residues S, N, D, F, C, K, G, and P) were not tolerated and decreased binding by more than 100-fold. Comparably stringent requirements were observed for position 9, where V, L and I were preferred and A and M are tolerated, while the residues T, C, N, F, and Y virtually abolished binding. According to this set of peptides, an optimal 2-9 motif could be defined with L, M in position 2 and V, I, or L in position 9.

TABLE 4

A2.1:BINDING OF ANALOGS OF A MOTIF-CONTAINING POLY A PEPTIDE

	pos. 1 A	pos. 2 L	pos. 3 A	pos. 4 K	pos. 5 A	pos. 6 A	pos. 7 A	pos. 8 A	pos. 9 V
A	1.00	0.013	1.00		1.00	1.00	1.00	1.00	0.070
G	0.46	*			0.63	0.12		0.57	
D	-	-	0.93	0.74	0.51	0.10			
E	0.012		0.68	1.53	0.62	0.15	0.28	0.26	
R						0.080			
H								0.24	
K	0.54	-	0.062	1.00	0.39		0.50	0.11	
L		1.00	0.46		0.99		0.76	0.90	0.11
V	0.47	0.051	0.15	1.12		0.44	0.49	0.30	1.00
I	0.41	0.063				1.12			0.18
M		0.43	0.66						0.024
Y	0.75		0.62		0.94	0.41	1.40	0.43	-
F	1.10	-	0.95			1.76		0.49	-
W									
Q					0.32		0.19	0.41	
N		-	0.34		1.24		0.97	0.31	-
S	0.44	-	0.37	0.97					
T	0.26	0.011		0.98			0.28	0.37	-
C		-		1.53		0.84			-
P	-	-	0.25	1.07		0.84	0.63	0.55	

Ratio ≤ 0.1

* Ratio ≤ 0.01

Non-Anchor Positions 1 and 3-8 in poly-A Analogs All non-anchor positions were more permissive to different substitutions than the anchor-positions 2 and 9, i.e most residues were tolerated. Significant decreases in binding were observed for some substitutions in distinct positions. More specifically, in position 1 a negative charge (residues D and E) or a P greatly reduced the binding capacity. Most substitutions were tolerated in position 3 with the exception of the residue K. Significant decreases were also seen in position 6 upon introduction of either a negative charge (D, E) or a positively charged residue (R). A summary of these effects by different single amino acid substitutions is given in Table 5.

TABLE 5

Summary A2.1 Poly-A

AA position	(+)	(+/-)	(-)
1	FAYKVGSIT		EDP
2	LM	VITA	SNDFCKGP
3	AFDEMYLSNPV	K	
4	CEVPATSD		
5	NALYGEDKQ		
6	FIAPCVYEG	DR	
7	YANLPVETQ		
8	ALGPFYQTNVEHK		
9	VIL	AM	TCNFY
	Ratio > 0.1	Ratio 0.01-0.1	Ratio < 0.01

The Motif-Library Approach To further evaluate the importance of non-anchor positions for binding, peptides of potential target molecules of viral and tumor origin were scanned for the presence of sequences containing optimal 2-9 anchor motifs. A set of 161 peptides containing a L or M in position 2 and a V, L or I in position 9 was selected, synthesized and tested for binding (see Example 6). Only 11.8% of these peptides bind with high affinity (ratio ≥0.10; 22.4% were intermediate binders (ratio ≥0.1). As many as 36% were weak binders (ratio <0.01 - 0.0001), and 31% were non-binders (ratio<0.0001). The high number of non-binders containing optimal anchor-motifs indicates that in this set of peptides positions other than the 2-9 anchors influence A2.1 binding capacity. Appendix 1 sets forth all of the peptides having the 2-9 motif used for this analysis and the binding data for those peptides.

To define the influence on non-anchor positions more specifically, the frequency of occurrence of each amino acid

in each of the non-anchor positions was calculated for the good and intermediate binders on one hand and non-binders on the other hand. Amino acids of similar chemical characteristic were grouped together. Weak binders were not considered for the following analysis. The frequency of occurence of each amino acid in each of the non-anchor positions was calculated for the good binders and non-binders (Table 6).

Several striking trends become apparent. For example in position 1, only 3.6% of the A2.1 binders and as much as 35% of the non-binders carried a negative charge (residues D and E). This observation correlates well with previous findings in the set of poly-A analogs, where a D or E substitution greatly affected binding. Similarly, the residue P was 8 times more frequent in non-binders than in good binders. Conversely, the frequencies of aromatic residues (Y, F, W) were greatly increased in A2.1 binders as compared to non-binders.

TABLE 6

A.2.1 9-mer PEPTIDES

NUMBER OF PEPTIDES	161	
GOOD BINDERS	19	11.8%
INTERMEDIATE BINDERS	36	22.4%
WEAK BINDERS	58	36.0%
NON-BINDERS	48	29.8%

	1+	1-	2+	2-	3+	3-	4+	4-	5+	5-	6+	6-	7+	7-	8+	8-	9+	9-
A	5.5	2.1	0.0	0.0	3.6	4.2	5.6	8.3	5.5	8.3	5.5	6.3	9.1	2.1	3.6	12.5	0.0	0.0
G	7.3	2.1	0.0	0.0	3.6	8.3	9.1	8.3	9.1	8.3	10.9	8.3	5.5	12.5	3.6	8.3	0.0	0.0
D,E	3.6	35.4	0.0	0.0	0.0	12.5	10.9	16.7	3.6	12.5	5.5	8.3	1.8	16.7	9.1	10.4	0.0	0.0
R,H,K	12.7	4.2	0.0	0.0	3.6	16.7	16.4	16.7	9.1	10.4	1.8	20.8	0.0	10.4	16.4	12.5	0.0	0.0
L,V,I,M	38.2	12.5	100.0	100.0	34.5	18.8	9.1	16.7	25.5	29.2	30.9	22.9	30.9	25.0	32.7	18.8	100.0	100.0
Y,F,W	14.5	2.1	0.0	0.0	21.8	4.2	7.3	8.3	18.2	2.1	16.4	8.3	14.5	8.3	5.5	8.3	0.0	0.0
Q,N	7.3	14.8	0.0	0.0	5.5	14.6	12.7	10.4	9.1	10.4	10.9	10.4	5.5	8.3	5.5	16.7	0.0	0.0
S,T,C	9.1	12.5	0.0	0.0	20.0	10.4	20.0	4.2	14.5	16.7	14.5	12.5	14.5	12.5	20.0	18.8	0.0	0.0
P	1.8	14.6	0.0	0.0	7.3	10.4	9.1	12.5	5.5	2.1	3.6	2.1	18.2	6.3	3.6	0.0	0.0	0.0
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Following this approach, amino acids of similar structural characteristics were grouped together. Then, the frequency of each amino acid group in each position was calculated for binders versus non-binders (Table 7). Finally, the frequency in the binders group was divided by the frequency in the non-binders to obtain a "frequency ratio". This ratio indicates whether a given amino-acid or group of residues occurs in a given position preferentially in good binders (ratio >1) or in non-binders (ratio <1).

TABLE 7

A2.1 9-mer PEPTIDES

NUMBER OF PEPTIDES	161	
GOOD BINDERS	19	11.8%
INTERMEDIATE BINDERS	36	22.4%
WEAK BINDERS	58	36.0%
NON-BINDERS	48	29.8%

	pos. 1 ratio	pos. 2 ratio	pos. 3 ratio	pos. 4 ratio	pos. 5 ratio	pos. 6 ratio	pos. 7 ratio	pos. 8 ratio	pos. 9 ratio
A	2.6	NA	0.9	0.9	0.7	0.9	4.4	0.3	NA
G	3.5	NA	0.4	1.1	1.1	1.3	0.4	0.4	NA
D,E	0.1	NA	0.0	0.7	0.3	0.7	0.1	0.9	NA
R,H,K	3.1	NA	0.2	1.0	0.9	0.1	0.0	1.3	NA
L,V,I,M	3.1	1.0	1.8	0.5	0.9	1.3	1.2	1.7	1.0
Y,F,W	7.0	NA	5.2	0.9	8.7	2.0	2.3	2.6	NA
Q,N	0.5	NA	0.4	1.2	0.9	1.0	0.7	0.3	NA
S,T,C	0.7	NA	1.9	4.8	0.9	1.2	1.2	1.1	NA
P	0.1	NA	0.7	0.7	2.6	1.7	2.9	+++	NA

+++ indicates that there were no negative binders

Different Residues Influence A2.1 Binding In order to analyse the most striking influences of certain residues on A2.1 binding, a threshold level was set for the ratios described in Table 7. Residues showing a more than 4-fold greater frequency in good binders were regarded as preferred residues (+). Residues showing a 4-fold lower frequency in A2.1 binders than in non-binders were regarded as disfavored residues (-). Following this approach, residues showing the most prominent positive or negative effects on binding are listed in Table 8.

This table identifies the amino acid groups which influence binding most significantly in each of the non-anchor positions. In general, the most negative effects were observed with charged amino acids. In position 1, negatively, charged amino acids were not observed in good binders, i.e., those amino acids were negative binding residues at position 1. The opposite was true for position 6 where only basic amino acids were detrimental for binding i.e., were negative binding residues. Moreover, both acidic and basic amino acids were not observed in A2.1 binders in positions 3 and 7. A greater than 4-fold increased frequency of non-binders was found when P was in position 1.

TABLE 8

Summary of A2.1 Motif-Library, 9-mers

AA POSITION	(+)	(-)
1	(YFW)	P, (DE)
2	Anchor	
3	(YFW)	(DE), (RKH)
4	(STC)	
5	(YFW)	
6		(RKH)
7	A	(RKH), (DE)
8		
9	Anchor	

(+) = Ratio ≥ 4-fold (-) = Ratio ≤ 0.25

Aromatic residues were in general favored in several of the non-anchor positions, particularly in positions 1, 3, and 5. Small residues like S, T, and C were favored in position 4 and A was favored in position 7.

An Improved A2.1 9-mer Motif The data described above was used to derive a stringent A2.1 motif. This motif is based in significant part on the effects of the non-anchor positions 1 and 3-8. The uneven distribution of amino acids at different positions is reflective of specific dominant negative binding effects of certain residues, mainly charged

ones, on binding affinity. A series of rules were derived to identify appropriate anchor residues in positions 2 and 9 and negative binding residues at positions 1 and 3-8 to enable selection of a high affinity binding immunogenic peptide. These rules are summarized in Table 9.

To validate the motif defined above and shown in Table 9 published sequences of peptides that have been naturally processed and presented by A2.1 molecules were analysed (Table 10). Only 9-mer peptides containing the 2-9 anchor residues were considered.

When the frequencies of these peptides were analysed, it was found that in general they followed the rules summarized in Table 9. More specifically, neither acidic amino acids nor P were found in position 1. Only one acidic amino acid and no basic amino acids were found in position 3. Positions 6 and 7 showed no charged residues. Acidic amino acids, however, were frequently found in position 8, where they are tolerated, according to our definition of the A2.1 motif. The analysis of the sequences of naturally processed peptides therefore reveals that >90% of the peptides followed the defined rules for a complete motif.

Thus the data confirms a role of positions other than the anchor positions 2 and 9 for A2.1 binding. Most of the deleterious effects on binding are induced by charged amino acids in non-anchor positions, i.e. negative binding residues occupying positions 1, 3, 6 or 7.

TABLE 9

A2.1 MOTIF FOR 9-MER PEPTIDES

AA Position	(+)	(-)
1		acidic amino-acids and P
2	Anchor: L, M, (I,V,A,T)	
3		acidic and basic amino-acids
4		
5		
6		basic amino-acids
7		acidic and basic amino-acids
8		
9	Anchor: V, I, L (A,M)	

TABLE 10

A2.1 naturally processed peptides

1	2	3	4	5	6	7	8	9	A2.1 binding
A	L	X	G	G	X	V	N	V	ND
L	L	D	V	P	T	A	A	V	ND
G	X	V	P	F	X	V	S	V	0.41
S	L	L	P	A	I	V	E	L	0.19
S	X	X	V	R	A	X	E	V	ND
Y	M	N	G	T	M	S	Q	V	ND
K	X	N	E	P	V	X	X	X	ND
Y	L	L	P	A	I	V	H	I	0.26
A	X	W	G	F	F	P	V	X	ND
T	L	W	V	D	P	Y	E	V	0.23
G	X	V	P	F	X	V	S	V	0.41

A2.1 Binding of Peptide 10-mers

The "Motif-Library" Approach Previous data clearly indicated that 10-mers can also bind to HLA molecules even if with a somewhat lower affinity than 9-mers. For this reason we expanded our analysis to 10-mer peptides.

Therefore, a "Motif-Library" set of 170 peptide 10-mers containing optimal motif-combinations was selected from known target molecule sequences of viral and tumor origin and analysed as described above for 9-mers. In this set we found 5.9% good binders, 17.1% intermediate binders, 41.2% weak binders and 35.9% non-binders. The actual sequences, origin and binding capacities of this set of peptides are included as Appendix 2. This set of 10-mers was used to determine a) the rules for 10-mer peptide binding to A2.1, b) the similarities or differences to rules defined for 9-mers, and c) if an insertion point can be identified that would allow for a superimposable common motif for 9-mers and 10-mers.

Amino-acid frequencies and frequency ratios for the various amino-acid groups for each position were generated for 10-mer peptides as described above for 9-mer peptides and are also shown in Tables 11 and 12, respectively for grouped residues.

A summary of preferred versus disfavored residues and of the rules derived for the 10-mers in a manner analogous to that used for 9-mers, is also listed in Tables 13 and 14, respectively.

When the frequency-ratios of different amino-acid groups in binders and non-binders at different positions were analysed and compared to the corresponding ratios for the 9-mers, both striking similarities and significant differences emerged (Table 15). At the N-terminus and the C-termini of 9-mers and 10-mers, similarities predominate. In position 1 for example, in 10-mers again the P residue and acidic amino acids were not tolerated. In addition at position 1 in 10-mers aromatic residues were frequently observed in A2.1 binders. In position 3, acidic amino acids were frequently associated with poor binding capacity in both 9-mers and 10-mers. Interestingly, however, while in position 3 aromatic residues

TABLE 11

A2.1 10-mer Peptides

NUMBER OF PEPTIDES	170		WEAK BINDERS	70 41.2%
GOOD BINDERS	10	5.9%	NON-BINDERS	61 35.9%
INTERMEDIATE BINDERS	29	17.1%		

	1+	1-	2+	2-	3+	3-	4+	4-	5+	5-
A	2.6	0.0	0.0	0.0	10.3	3.3	2.6	11.5	5.1	3.3
G	7.7	9.8	0.0	0.0	7.7	16.4	15.4	3.3	5.1	6.6
D,E	0.0	23.0	0.0	0.0	2.6	16.4	7.7	13.1	2.6	9.8
R,H,K	7.7	6.6	0.0	0.0	5.1	16.4	2.8	16.0	10.3	14.8
L,V,I,M	48.7	16.4	100.0	100.0	33.3	3.3	23.1	23.0	30.8	24.6
Y,F,W	12.8	0.0	0.0	0.0	12.6	4.9	15.4	4.9	17.9	4.9
Q,N	10.3	9.6	0.0	0.0	7.7	8.2	7.7	9.8	7.7	9.8
S,T,C	10.3	11.5	0.0	0.0	15.4	18.0	12.6	11.5	20.5	19.7
P	0.0	23.0	0.0	0.0	5.1	13.1	12.6	4.9	0.0	6.6
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

	6+	6-	7+	7-	8+	8-	9+	9-	10+	10-
A	7.7	13.1	10.3	8.2	7.7	4.9	2.6	4.9	0.0	0.0
G	10.3	1.6	17.9	6.6	7.7	11.5	7.7	9.8	0.0	0.0
D,E	10.3	9.6	5.1	15.4	0.0	16.4	5.1	13.1	0.0	0.0
R,H,K	7.7	19.7	2.6	14.8	0.0	29.5	2.6	16.4	0.0	0.0
L,V,I,M	30.8	14.5	25.5	18.0	23.1	4.9	12.8	16.4	100.0	100.0
Y,F,W	7.7	13.1	12.8	6.2	23.1	1.6	20.5	9.6	0.0	0.0
Q,N	2.6	3.3	5.1	8.2	2.6	6.6	7.7	11.5	0.0	0.0
S,T,C	17.9	19.7	17.9	13.1	20.5	16.4	33.3	11.5	0.0	0.0
P	5.1	4.9	2.6	5.6	15.4	8.2	7.7	6.6	0.0	0.0
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

TABLE 12

A2.1 10-mer Peptides

NUMBER OF PEPTIDES	170	
GOOD BINDERS	10	5.9%
INTERMEDIATE BINDERS	29	17.1%
WEAK BINDERS	70	41.2%
NON-BINDERS	61	35.9%

	pos. 1 ratio	pos. 2 ratio	pos. 3 ratio	pos. 4 ratio	pos. 5 ratio	pos. 6 ratio	pos. 7 ratio	pos. 8 ratio	pos. 9 ratio	pos. 10 ratio
A	+++	NA	3.1	0.2	1.8	0.6	1.3	1.6	0.5	NA
G	0.8	NA	0.5	4.7	0.8	6.3	2.7	0.7	0.8	NA
D,E	0.0	NA	0.2	0.6	0.3	1.0	0.3	0.0	0.4	NA
R,H,K	1.2	NA	0.3	0.1	0.7	0.4	0.2	0.0	0.2	NA
L,V,I,M	3.0	1.0	10.2	1.0	1.3	2.1	1.4	4.7	0.8	1.0
Y,F,W	+++	NA	2.6	3.1	3.6	0.6	1.6	14.1	2.1	NA
Q,N	1.0	NA	0.9	0.8	0.8	0.8	0.6	0.4	0.7	NA
S,T,C	0.9	NA	0.9	1.1	1.0	0.9	1.4	1.3	2.9	NA
P	0.0	NA	0.4	2.6	0.0	1.0	0.4	1.9	1.2	NA

+++ Indicates that there were no negative binders.

TABLE 13
Summary of A2.1 Motif-Library 10-mers

AA position	(+)	(-)
1	(YFW), A	(DE), P
2	Anchor	
3	(LVIM)	(DE)
4	G	A, (RKH)
5		P
6	G	
7		(RKH)
8	(YFW), (LVIM)	(DE), (RKH)
9		(RKH)
10	Anchor	

(+) = Ratio ≥ 4-fold (-) = Ratio ≤ 0.25

TABLE 14

A2.1 MOTIF FOR 10-MER PEPTIDES

AA Position	(+)	(-)
1		acidic amino-acids and P
2	Anchor: L, M, (I,V,A,T)	
3		acidic amino-acids
4		basic amino-acids and A
5		P
6		
7		basic amino-acids
8		acidic and basic amino-acids
9		basic amino-acids
10	Anchor: V, I, L (A,M)	

TABLE 15

COMPARISON OF A2.1 BINDING OF 9-MERS AND 10-MERS

AA Position	9-mers (+)	10-mers (+)
1	(YFW)	(YFW)
2	Anchor	Anchor
3	(YWF)	(LVIM)
4	(STC)	G
5	(YWF)	
6		G
7	A	
8		(YWF), (LVIM)
9	Anchor	
10	-	Anchor

AA Position	9-mers (-)	10-mers (-)
1	P, (DE)	P, (DE)
2	Anchor	Anchor
3	(DE), (RKH)	(DE)
4		A, (RKH)
5		P
6	(RKH)	
7	(DE), (RKH)	(RKH)
8		(DE), (RKH)
9	Anchor	(RKH)
10	-	Anchor

were preferred in 9-mers, aliphatic residues (L, V, I, M) were preferred in 10-mers.

At the C-terminus of the peptides, basic amino acids are not favored in position 7, and both acidic and basic amino acids are not favored in position 8 for 10-mers. This is in striking agreement with the observation that the same pattern was found in 9-mers at positions 6 and 7. Interestingly, again the favored residues differ between two peptides sizes. Aromatic (Y, F, W) or aliphatic (L, V, I, M) residues were preferred in 10-mers at position 8, while the A residue was preferred by 9-mers in the corresponding position 7.

By contrast, in the center of the peptide no similarities of frequency preferences were observed at positions 4, 5, and 6 in 10-mers and positions 4 and 5 in the 9-mers.

Most interestingly, among the residues most favored in the center of the tested peptides were G in position 4 and 6, P in position 5 was not observed in binders. All of these residues are known to dramatically influence the overall secondary structure of peptides, and in particular would be predicted to strongly influence the propensity of a 10-mer to adopt a "kinked" or "bulged" conformation.

Charged residues are predominantly deleterious for binding and are frequently observed in non-binders of 9-mers and 10-mers.

However, favored residues are different for 9-mers and 10-mers. Glycine is favored while Proline is disfavored in the center of 10-mer peptides but this is not the case for 9-mers.

These data establish the existence of an "insertion area" spanning two positions (4, 5) in 9-mers and 3 positions (4, 5, 6) in 10-mers. This insertion area is a more permissive region where few residue similarities are observed between the 9-mer and 10-mer antigenic peptides. Furthermore, in addition to the highly conserved anchor positions 2 and 9, there are "anchor areas" for unfavored residues in positions 1 and 3 at the N-terminus for both 9-mer and 10-mer and

positions 7-10 or 6-9 at the C-terminus for 10-mers and 9-mers, respectively.

Example 6

Algorithm to Predict Binding of 9-mer Peptides to HLA-A2.1

Within the population of potential A2.1 binding peptides identified by the 2,9 motif, as shown in the previous example, only a few peptides are actually good or intermediate binders and thus potentially immunogenic. It is apparent from the data previously described that the residues present in positions other than 2 and 9 can influence, often profoundly, the binding affinity of a peptide. For example, acidic residues at position 1 for A2.1 peptides do not appear to be tolerated. Therefore, a more exact predictor of binding could be generated by taking into account the effects of different residues at each position of a peptide sequence, in addition to positions 2 and 9.

More specifically, we have utilized the data bank obtained during the screening of our collection of A2.1 motif containing 9-mer peptides to develop an algorithm which assigns a score for each amino acid, at each position along a peptide. The score for each residue is taken as the ratio of the frequency of that residue in good and intermediate binders to the frequency of occurrence of that residue in non-binders.

In the present "Grouped Ratio" algorithm residues have been grouped by similarity. This avoids the problem encountered with some rare residues, such as tryptophan, where there are too few occurrences to obtain a statistically significant ratio. Table 16 is a listing of scores obtained by grouping for each of the twenty amino acids by position for 9-mer peptides containing perfect 2/9 motifs. A peptide is scored in the "Grouped Ratio" algorithm as a product of the scores of each of its residues. In the case of positions other than 2 and 9, the scores have been derived using a set of peptides which contain only preferred residues in positions 2 and 9. To enable us to extend our "Grouped Ratio" algorithm

TABLE 16

	1	2	3	4	5	6	7	8	9
A	2.6	0.03	0.87	0.87	0.65	0.87	4.4	0.29	0.16
C	0.73	0.01	1.9	4.8	0.87	1.2	1.2	1.1	0.01
D	0.10	0.01	0.10	0.65	0.29	0.65	0.11	0.87	0.01
E	0.10	0.01	0.10	0.65	0.29	0.65	0.11	0.87	0.01
F	7.0	0.01	5.2	0.87	8.7	2.0	2.3	2.6	0.01
G	3.5	0.01	0.44	1.1	1.1	1.3	0.44	0.44	0.01
H	3.1	0.01	0.22	1.0	0.87	0.09	0.10	1.3	0.01
I	3.1	0.14	1.8	0.55	0.87	1.4	1.2	1.8	0.40
K	3.1	0.01	0.22	1.0	0.87	0.09	0.10	1.3	0.01
L	3.1	1.00	1.8	0.55	0.87	1.4	1.2	1.8	0.09
M	3.1	2.00	1.8	0.55	0.87	1.4	1.2	1.8	0.06
N	0.50	0.01	0.37	1.2	0.87	1.1	0.65	0.33	0.01
P	0.12	0.01	0.70	0.73	2.6	1.8	2.9	0.10	0.01
Q	0.50	0.01	0.37	1.2	0.87	1.1	0.65	0.33	0.01
R	3.1	0.01	0.22	1.0	0.87	0.09	0.10	1.3	0.01
S	0.73	0.01	1.9	4.8	0.87	1.2	1.2	1.1	0.01
T	0.73	0.01	1.9	4.8	0.87	1.2	1.2	1.1	0.01
V	3.1	0.08	1.8	0.55	0.87	1.4	1.2	1.8	1.00
W	7.0	0.01	5.2	0.87	8.7	2.0	2.3	2.6	0.01
Y	7.0	0.01	5.2	0.87	8.7	2.0	2.3	2.6	0.01

to peptides which may have residues other than the preferred ones at 2 and 9, scores for 2 and 9 have been derived from a set of peptides which are single amino acid substitutions at positions 2 and 9. Figure 2 shows a scattergram of the log of relative binding plotted against "Grouped Ratio" algorithm score for our collection of 9-mer peptides from the previous example.

The present "Grouped Ratio" algorithm can be used to predict a population of peptides with the highest occurrence of good binders. If one were to rely, for example, solely on a 2(L,M) and 9(V) motif for predicting A2.1 binding 9-mer peptides, it would have been predicted that all 160 peptides in our database would be good binders. In fact, as has already been described, only 12% of these peptides would be described as good binders and only 22% as intermediate binders; 66% of the peptides predicted by such a 2,9 motif are either weak or non-binding peptides. In contrast, using the "Grouped Ratio" algorithm described above, and selecting a score of 1.0 as threshold, 41 peptides were selected. Of this set, 27% are good binders, and 49% are intermediate, while only 20% are weak and 5% are non-binders (Table 17).

The present example of an algorithm has used the ratio of binders/non-binders to measure the impact of a particular residue at each position of a peptide. It is immediately apparent to one of ordinary skill that there are alternative ways of creating a similar algorithm.

An algorithm using the average binding affinity of all the peptides with a certain amino acid (or amino acid type) at a certain position has the advantage of including all of the peptides in the analysis, and not just good/intermediate binders and non-binders. Moreover, it gives a more quantitative measure of affinity than the simpler "Grouped Ratio" algorithm. We have created such an algorithm by calculating for each amino acid, by position, the average log of binding when that particular residue occurs in our set of 160 2,9 motif containing peptides. These values are shown in Table 18. The algorithm score for a peptide is then taken as the sum of the scores by position for each residues.

Figure 3 shows a scattergram of the log of relative binding against the average "Log of Binding" algorithm score. Table 17 shows the ability of the two algorithms to predict peptide binding at various levels, as a function of the cut-off score used. The ability of a 2,9 motif to predict binding in the same peptide set is also shown for reference purposes. It is clear from this comparison that both algorithms of this invention have a greater ability to predict populations with higher frequencies of good binders than a 2,9 motif alone. Differences between the "Grouped Ratio" algorithm and the "Log of Binding" algorithm are small in the set of peptides analyzed here, but do suggest that the "Log of Binding" algorithm is a better, if only slightly, predictor than the "Grouped Ratio" algorithm.

The log of binding algorithm was further revised in two ways. First, poly-alanine (poly-A) data were incorporated into the algorithms at the anchor positions for residues included in the expanded motifs where data obtained by screening a large library of peptides were not available. Second, an "anchor requirement screening filter" was incorporated into the algorithm. The poly-A approach is described in detail, above. The "anchor requirement screening filter" refers to the way in which residues are scored at the anchor positions, thereby providing the ability to screen out peptides which do not have preferred or tolerated residues in the anchor positions. This is accomplished by assigning a score for unacceptable residues at the anchor positions which are so high as to preclude any peptide which contains them from achieving an overall score which would allow it to be considered as a potential binder.

The results for 9-mers and 10-mers are presented in Tables 26 and 27, below. In these tables, values are group values as follows: A; G; P; D,E; R,H,K; L,I,V,M; F,Y,W; S,T,C; and Q,N, except where noted in the tables.

TABLE 17

Criteria	Cut-off	Good Binders	Intermediate Binders	Weak Binders	Negative Binders	Totals
2.9 motif		19 (12%)	36 (22%)	58 (36%)	48 (30%)	161 (100%)
Grouped Ratio Algorithm	1.5	5 (83%)	1 (17%)	0 (0%)	0 (0%)	6 (100%)
	1.25	8 (67%)	4 (33%)	0 (0%)	0 (0%)	12 (100%)
	1	10 (50%)	9 (45%)	1 (5%)	0 (0%)	20 (100%)
	0.5	12 (32%)	17 (46%)	7 (19%)	1 (3%)	37 (100%)
	0	12 (23%)	26 (49%)	12 (23%)	3 (6%)	53 (100%)
	-1	17 (18%)	35 (37%)	33 (35%)	10 (11%)	95 (100%)
	-2	19 (15%)	36 (29%)	50 (40%)	21 (17%)	126 (100%)
	-3	19 (13%)	36 (24%)	56 (38%)	38 (26%)	149 (100%)
	no cut	19 (12%)	36 (22%)	58 (36%)	48 (30%)	161 (100%)
Log of Binding Algorithm	-19	5 (100%)	0 (0%)	0 (0%)	0 (0%)	5 (100%)
	-20	8 (73%)	3 (27%)	0 (0%)	0 (0%)	11 (100%)
	-21	15 (43%)	15 (43%)	5 (14%)	0 (0%)	35 (100%)
	-22	17 (26%)	27 (41%)	21 (32%)	1 (2%)	68 (100%)
	-23	18 (19%)	35 (37%)	34 (36%)	7 (7%)	94 (100%)
	-24	18 (16%)	36 (30%)	47 (39%)	17 (14%)	119 (100%)
	-25	19 (14%)	36 (26%)	55 (39%)	30 (21%)	140 (100%)
	no cut	19 (12%)	36 (22%)	58 (36%)	48 (30%)	161 (100%)

TABLE 18

	1	2	3	4	5	6	7	8	9
A	-2.38	-3.22	-2.80	-2.68	-2.89	-2.70	-2.35	-3.07	-2.49
C	-2.94	-4.00	-2.58	-1.96	-3.29	-2.22	-2.97	-2.37	-4.00
D	-3.69	-4.00	-3.46	-2.71	-2.26	-2.63	-3.61	-3.03	-4.00
E	-3.64	-4.00	-3.51	-2.65	-3.39	-3.41	-3.21	-2.63	-4.00
F	-1.89	-4.00	-2.35	-2.50	-1.34	-2.43	-2.18	-1.71	-4.00
G	-2.32	-4.00	-3.04	-2.63	-2.56	-2.30	-3.13	-2.96	-4.00
H	-2.67	-4.00	-2.58	-2.58	-2.05	-3.32	-3.13	-2.16	-4.00
I	-1.65	-2.55	-2.80	-3.44	-2.74	-2.79	-2.20	-2.69	-2.10
K	-2.51	-4.00	-3.65	-2.93	-3.34	-3.77	-3.13	-3.27	-4.00
L	-2.32	-1.70	-2.02	-2.49	-2.71	-2.63	-2.62	-2.01	-2.74
M	-0.39	-1.39	-1.79	-3.07	-3.43	-1.38	-1.33	-0.97	-2.96
N	-3.12	-4.00	-3.52	-2.22	-2.36	-2.30	-3.14	-3.31	-4.00
P	-3.61	-4.00	-2.97	-2.64	-2.42	-2.31	-1.83	-2.42	-4.00
Q	-2.76	-4.00	-2.81	-2.63	-3.06	-2.84	-2.12	-3.05	-4.00
R	-1.92	-4.00	-3.41	-2.61	-3.05	-3.76	-3.43	-3.02	-4.00
S	-2.39	-3.52	-2.04	-2.12	-2.83	-3.04	-2.73	-2.02	-4.00
T	-2.92	-4.00	-2.60	-2.48	-2.17	-2.58	-2.67	-3.14	-3.70
V	-2.44	-2.64	-2.68	-3.29	-2.49	-2.24	-2.68	-2.83	-1.70
W	-0.14	-4.00	-1.01	-2.94	-1.63	-2.77	-2.85	-2.13	-4.00
X	-1.99	-2.13	-2.41	-2.97	-2.72	-2.70	-2.41	-2.35	-2.42
Y	-1.46	-4.00	-1.67	-2.70	-1.92	-2.39	-1.35	-3.37	-4.00

Example 7

Use of an Algorithm to Predict Binding of 10-mer Peptides to HLA-A2.1

Using the methods described in the proceeding example, an analogous set of algorithms has been developed for predicting the binding of 10-mer peptides. Table 19 shows the scores used in a "Grouped Ratio" algorithm, and Table 20 shows the "Log of Binding" algorithm scores, for 10-mer peptides. Table 21 shows a comparison of the application of the two different algorithmic methods for selecting binding peptides. Figures 4 and 5 show, respectively, scattergrams of a set of 10-mer peptides containing preferred residues in positions 2 and 10 as scored by the "Grouped Ratio" and "Log of Binding" algorithms.

TABLE 19

	1	2	3	4	5	6	7	8	9	10
A	3.00	0.01	3.10	0.20	1.60	0.60	1.30	1.60	0.50	0.01
C	0.90	0.01	0.90	1.10	1.00	0.90	1.40	1.30	2.90	0.01
D	0.01	0.01	0.20	0.60	0.30	1.00	0.30	0.01	0.40	0.01
E	0.01	0.01	0.20	0.60	0.30	1.00	0.30	0.01	0.40	0.01
F	3.00	0.01	2.60	3.10	3.60	0.60	1.60	14.1	2.10	0.01
G	0.80	0.01	0.50	4.70	0.80	6.30	2.70	0.70	0.80	0.01
H	1.20	0.01	0.30	0.10	0.70	0.40	0.20	0.01	0.20	0.01
I	3.00	0.50	10.2	1.00	1.30	2.10	1.40	4.70	0.80	1.00
K	1.20	0.01	0.30	0.10	0.70	0.40	0.20	0.01	0.20	0.01
L	3.00	1.10	10.2	1.00	1.30	2.10	1.40	4.70	0.80	0.50
M	3.00	0.60	10.2	1.00	1.30	2.10	1.40	4.70	0.80	0.01
N	1.00	0.01	0.90	0.80	0.80	0.80	0.60	0.40	0.70	0.01
P	0.00	0.01	0.40	2.60	0.01	1.00	0.40	1.90	1.20	0.01
Q	1.00	0.01	0.90	0.80	0.80	0.80	0.60	0.40	0.70	0.01
R	1.20	0.01	0.30	0.10	0.70	0.40	0.20	0.01	0.20	0.01
S	0.90	0.01	0.90	1.10	1.00	0.90	1.40	1.30	2.90	0.01
T	0.90	0.01	0.90	1.10	1.00	0.90	1.40	1.30	2.90	0.01
V	3.00	0.10	10.2	1.00	1.30	2.10	1.40	4.70	0.80	2.30
W	3.00	0.01	2.60	3.10	3.60	0.60	1.60	14.1	2.10	0.01
Y	3.00	0.01	2.60	3.10	3.60	0.60	1.60	14.1	2.10	0.01

TABLE 20

	1	2	3	4	5	6	7	8	9	10	
A	-2.40	-4.00	-2.54	-3.42	-3.07	-3.30	-2.98	-2.69	-3.29	-4.00	
C	-3.64	-4.00	-2.47	-2.48	-1.78	-3.94	-1.28	-3.10	-2.43	-4.00	
D	-3.65	-4.00	-2.76	-3.26	-2.76	-3.03	-3.43	-3.68	-3.63	-4.00	
E	-3.92	-4.00	-3.63	-3.34	-3.73	-2.82	-3.54	-3.71	-2.95	-4.00	
F	-1.52	-4.00	-1.96	-3.03	-2.01	-3.11	-2.67	-1.61	-2.43	-4.00	
G	-2.91	-4.00	-3.40	-2.63	-2.98	-2.45	-2.52	-3.18	-3.03	-4.00	
H	-3.61	-4.00	-3.10	-3.03	-2.33	-2.99	-3.70	-3.55	-4.00	-4.00	
I	-2.26	-4.00	-2.82	-3.05	-2.38	-2.61	-2.38	-3.34	-3.18	-1.47	
K	-2.53	-4.00	-3.65	-3.42	-3.14	-3.58	-3.50	-3.53	-4.00	-4.00	
L	-2.00	-2.93	-2.21	-2.48	-2.88	-2.53	-2.57	-1.83	-3.23	-3.20	
M	-2.41	-3.11	-2.00	-3.33	-3.70	-2.56	-3.27	-2.25	-3.00	-4.00	
N	-3.21	-4.00	-3.09	-2.61	-2.93	-2.89	-3.52	-3.01	-2.88	-4.00	
P	-3.90	-4.00	-3.21	-2.27	-3.72	-3.06	-3.35	-2.58	-2.94	-4.00	
Q	-2.92	-4.00	-2.97	-4.00	-2.98	-3.46	-2.20	-3.23	-3.45	-4.00	
R	-3.01	-4.00	-3.44	-3.50	-3.23	-3.32	-3.72	-3.59	-2.97	-4.00	
S	-2.47	-4.00	-3.17	-3.11	-3.23	-2.64	-3.19	-2.79	-2.26	-4.00	
T	-3.59	-4.00	-3.07	-2.88	-2.89	-3.16	-2.43	-3.11	-2.58	-4.00	
V	-2.97	-4.00	-2.46	-3.14	-3.27	-2.53	-3.14	-3.02	-2.90	-2.61	
W	-2.10	-4.00	-2.72	-1.79	-2.65	-1.92	-1.80	-2.24	-2.11	-4.00	
Y	-2.37	-4.00	-2.42	-2.85	-3.03	-3.76	-2.82	-2.34	-2.74	-4.00	

TABLE 21

Criteria	Cut-off	Good	Intermediate	Weak	Negative	Totals
2,10 motif		10 (6%)	29 (17%)	70 (41%)	61 (36%)	170 (100%)
Grouped Ratio Algorithm	4	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
	3	1 (25%)	2 (50%)	1 (25%)	0 (0%)	4 (100%)
	2	6 (24%)	13 (52%)	6 (24%)	0 (0%)	25 (100%)
	1	10 (21%)	21 (45%)	16 (34%)	0 (0%)	47 (100%)
	0	10 (15%)	28 (42%)	26 (39%)	2 (3%)	66 (100%)
	-1	10 (11%)	29 (32%)	42 (46%)	11 (12%)	92 (100%)
	-2	10 (9%)	29 (25%)	54 (47%)	23 (20%)	116 (100%)
	-3	10 (7%)	29 (22%)	63 (47%)	32 (24%)	134 (100%)
	no cut	10 (6%)	29 (17%)	70 (41%)	61 (36%)	170 (100%)
Log of Binding Algorithm	-24	2 (50%)	2 (50%)	0 (0%)	0 (0%)	4 (100%)
	-25	5 (56%)	3 (33%)	1 (11%)	0 (0%)	9 (100%)
	-26	7 (47%)	5 (33%)	3 (20%)	0 (0%)	15 (100%)
	-27	10 (32%)	9 (29%)	12 (39%)	0 (0%)	31 (100%)
	-28	10 (17%)	19 (33%)	29 (50%)	0 (0%)	58 (100%)
	-29	10 (12%)	25 (30%)	48 (58%)	0 (0%)	83 (100%)
	-30	10 (10%)	29 (28%)	59 (57%)	5 (5%)	103 (100%)
	-31	10 (8%)	28 (22%)	66 (51%)	24 (19%)	129 (100%)
	-32	10 (7%)	29 (19%)	70 (47%)	40 (27%)	149 (100%)
	no cut	10 (6%)	29 (17%)	70 (41%)	61 (36%)	170 (100%)

Example 8

Binding of A2.1 Algorithm Predicted Peptides

The results of Examples 6 and 7 indicate that an algorithm can be used to select peptides that bind to HLA-A2.1 sufficiently to have a high probability of being immunogenic.

To test this result, we tested our algorithm on a large (over 1300) non-redundant, independent set of peptides derived from various sources. After scoring this set with our algorithm, we selected 41 peptides (Table 21) for synthesis, and tested them for A2.1 binding. This set of peptides was comprised of 21 peptides with high algorithm scores, and 20 peptides with low algorithm scores.

The binding data and categorization profile are shown in Tables 22 and 23 respectively. The correlation between binding and algorithm score was 0.69. It is immediately apparent from Table 23 the striking difference between peptides with high algorithm scores, and those with low algorithm scores. Respectively, 76% of the high scorers and none of the low scorers were either good or intermediate binders. This data demonstrates the utility of the algorithm of this invention.

TABLE 22

SEQUENCE	SOURCE	A2.1 Binding	Algorithm Score
MMWFVVLTV	CMV	0.76	346
YLLLYFSPV	CMV	0.75	312
YLYRLNFCL	CMV	0.72	169
FMWTYLVTL	CMV	0.68	336
LLWWITILL	CMV	0.49	356
GLWCVLFFV	CMV	0.47	1989
LMIRGVLEV	CMV	0.45	296
LLLCRLPFL	CMV	0.42	1356
RLLTSLFFL	HSV	0.34	859
LLLYYDYSL	HSV	0.28	390
AMSRNLFRV	CMV	0.15	1746
AMLTACVEV	CMV	0.089	411
RLQPNVPLV	CMV	0.048	392
VLARTFTPV	CMV	0.044	196
RLLRGURL	CMV	0.037	494
WMWFPSVLL	CMV	0.036	362
YLCCGITLL	CMV	0.021	1043
DMLGRVFFV	HSV	0.011	1422
ALGRYQQLV	CMV	0.0089	184
LMPPPVAEL	CMV	0.0066	416
LMCRYTPRL	CMV	0.0055	414
RLTWRLTWL	CMV	0.0052	250
AMPRRVLHV	CMV	0.0014	628
ALLLVLALL	CMV	0.0014	535
AMSGTGTTL	CMV	0.0005	602
MLNVMKEAV	CMV	0.0039	0.00031
TMELMIRTV	CMV	0.0029	0.0013
TLAAMHSKL	HSV	0.0008	0.0019
TLNIVRDHV	CMV	0.0005	0.00021
ELSIFRERL	HSV	0.0002	0.0020
FLRVQQKAL	HSV	0.0002	0.00099
ELQMMQDWV	CMV	0.0001	0.0020
QLNAMKPDL	MT	0.0001	0.0017
GLRQLKGAL	CMV	0.0001	0.0010
TLRMSSKAV	HSV	0.0001	0.00085
SLRIKRELL	CMV	0	0.00041
DLKQMERVV	CMV	0	0.00026
PLRVTPSDL	CMV	0	0.0019
QLDYEKQVL	CMV	0	0.0012
WLKLLRDAL	CMV	0	0.0012
PMEAVRHPL	CMV	0	0.0011
ELKQTRVNL	CMV	0	0.00053
NLEVIHDAL	CMV	0	0.00050
ELKKVKSVL	HSV	0	0.00033
PLAYERDKL	CMV	0	0.00017

TABLE 23

Set	Good Binders	Intermediate Binders	Weak Binders	Negative Binders	Totals
HI Scorers	11 (52.4%)	5 (23.8%)	5 (23.8%)	0 (0.0%)	21 (100%)
Low Scorers	0 (0.0%)	C (0.0%)	10 (50.0%)	10 (50.0%)	20 (100%)
Totals	11 (26.6%)	5 (12.2%)	15 (36.6%)	10 (24.4%)	41 (100%)

Example 9

Ex vivo induction of Cytotoxic T Lymphocytes (CTL)

Peripheral blood mononuclear cells (PBMC) are isolated from an HLA-typed patient by either venipuncture or apheresis (depending upon the initial amount of CTLp required), and purified by gradient centrifugation using Ficoll-Paque (Pharmacia). Typically, one can obtain one million PBMC for every ml of peripheral blood, or alternatively, a typical apheresis procedure can yield up to a total of 1-10 X 10^{10} PBMC.

The isolated and purified PBMC are co-cultured with an appropriate number of antigen presenting cell (APC), previously incubated ("pulsed") with an appropriate amount of synthetic peptide (containing the HLA binding motif and the sequence of the antigen in question). PBMC are usually incubated at 1-2 X 10^6 cells/ml in culture medium such as RPMI-1640 (with autologous serum or plasma) or the serum-free medium AIM-V (Gibco).

APC are usually used at concentrations ranging from $1X10^4$ to $2X10^5$ cells/ml, depending on the type of cell used. Possible sources of APC include: 1) autologous dendritic cells (DC), which are isolated from PBMC and purified as described (Inaba, et al., J. Exp. Med. 166:182 (1987)); and 2) mutant and genetically engineered mammalian cells that express "empty" HLA molecules (which are syngeneic [genetically identical] to the patient's allelic HLA form), such as the, mouse RMA-S cell line or the human T2 cell line. APC containing empty HLA molecules are known to be potent inducers of CTL responses, possibly because the peptide can associate more readily with empty MHC molecules than with MHC molecules which are occupied by other peptides (DeBruijn, et al., Eur. J. Immunol. 21:2963-2970 (1991)).

In those cases when the APC used are not autologous, the cells will have to be gamma irradiated with an appropriate dose (using, e.g., radioactive cesium or cobalt) to prevent their proliferation both ex vivo, and when the cells are re-introduced into the patients.

The mixture cultures, containing PBMC, APC and peptide are kept in an appropriate culture vessel such as plastic T-flasks, gas-permeable plastic bags, or roller bottles, at 37° centigrade in a humid air/CO_2 incubator. After the activation phase of the culture, which usually occurs during the first 3-5 days, the resulting effector CTL can be further expanded, by the addition of recombinant DNA-derived growth factors such as interleukin-2 (IL-2), interleukin-4 (IL-4), or interleukin-7 (IL-7) to the cultures. An expansion culture can be kept for an additional 5 to 12 days, depending on the numbers of effector CTL required for a particular patient. In addition, expansion cultures may be performed using hollow fiber artificial capillary systems (Cellco), where larger numbers of cells (up to $1X10^{11}$) can be maintained.

Before the cells are infused into the patient, they are tested for activity, viability, toxicity and sterility. The cytotoxic activity of the resulting CTL can be determined by a standard ^{51}Cr-release assay (Biddison, W.E. 1991, Current Protocols in Immunology, p7,17.1-7.17.5, Ed. J. Coligan et al., J. Wiley and Sons, New York), using target cells that express the appropriate HLA molecule, in the presence and absence of the immunogenic peptide. Viability is determined by the exclusion of trypan blue dye by live cells. Cells are tested for the presence of endotoxin by conventional techniques. Finally, the presence of bacterial or fungal contamination is determined by appropriate microbiological methods (chocolate agar, etc.). Once the cells pass all quality control and safety tests, they are washed and placed in the appropriate infusion solution (Ringer/glucose lactate) and infused intravenously into the patient.

Example 10

Assays for CTL Activity

1. Peptide synthesis: Peptide syntheses were carried out by sequential coupling of N-α-Fmoc-protected amino acids on an Applied Biosystems (Foster City, CA) 430A peptide synthesizer using standard Fmoc coupling cycles (software

version 1.40). All amino acids, reagents, and resins were obtained from Applied Biosystems or Bachem. Solvents were obtained from Burdick & Jackson. Solid-phase synthesis was started from an appropriately substituted Fmoc-amino acid-Sasrin resin. The loading of the starting resin was 0.5-0.7 mmol/g polystyrene, and 0.1 or 0.25 meq were used in each synthesis. A typical reaction cycle proceeded as follows: 1) The N-terminal Fmoc group was removed with 25% piperidine in dimethylformamide (DMF) for 5 minutes, followed by another treatment with 25% piperdine in DMF for 15 minutes. The resin was washed 5 times with DMF. An N-methylpyrolidone (NMP) solution of a 4 to 10 fold excess of a pre-formed 1-hydroxybenzotriazole ester of the appropriate Fmoc-amino acid was added to the resin and the mixture was allowed to react for 30-90 min. The resin was washed with DMF in preparation for the next elongation cycle. The fully protected, resin bound peptide was subjected to a piperidine cycle to remove the terminal Fmoc group. The product was washed with dichloromethane and dried. The resin was then treated with trifluoroacetic acid in the presence of appropriate scavengers [e.g. 5% (v/v) water] for 60 minutes at 20°C. After evaporation of excess trifluoroacetic acid, the crude peptide was washed with dimethyl ether, dissolved in water and lyophilized. The peptides wee purified to >95% homogeneity by reverse-phase HPLC using H_2O/CH_3CN gradients containing 0.2% TFA modifier on a Vydac, 300Å pore-size, C-18 preparative column. The purity of the synthetic peptides was assayed on an analytical reverse-phase column, and their composition ascertained by amino acid analysis and/or sequencing. Peptides were routinely dissolved in DMSO at the concentration of 20 mg/ml.

2. Media: RPMI-1640 containing 10% fetal calf serum (FCS) 2 mM Glutamine, 50 μg/ml Gentamicin and $5x10^{-5}$M 2-mercaptoethanol served as culture medium and will be referred to as R10 medium.

RPMI-1640 containing 25 mM Hepes buffer and supplemented with 2% FCS was used as cell washing medium.

3. Rat Concanavalin A supernatant: The spleen cells obtained from Lewis rats (Sprague-Dawley) were resuspended at a concentration of $5x10^6$ cells/ml in R10 medium supplemented with 5 μg/ml of ConA in 75 cm2 tissue culture flasks. After 48 hr at 37°C, the supernatants were collected, supplemented with 1% α-methyl-D-mannoside and filter sterilized (.45 μm filter). Aliquots were stored frozen at -20°C.

4. LPS-activated lymphoblasts: Murine splenocytes were resuspended at a concentration of $1-1.5x10^6$/ml in R10 medium supplemented with 25 μg/ml LPS and 7 μg/ml dextran sulfate in 75 cm^2 tissue culture flasks. After 72 hours at 37°C, the lymphoblasts were collected for use by centrifugation.

5. Peptide coating of lymphoblasts: Coating of the LPS activated lymphoblasts was achieved by incubating $30x10^6$ lymphoblasts with 100 μg of peptide in 1 ml of R10 medium for 1 hr at 37°C. Cells were then washed once and resuspended in R10 medium at the desired concentration for use in in vitro CTL activation.

6. Peptide coating of Jurkat A2/K^b cells: Peptide coating was achieved by incubating $10x10^6$ irradiated (20,000 rads) Jurkat A2.1/K^b cells with 20 μg of peptide in 1 ml of R10 medium for 1 hour at 37°C. Cells were washed three times and resuspended at the required concentration in R10 medium.

7. In Vitro CTL activation: One to four weeks after priming spleen cells (5 x 10^6 cells/well or $30x10^6$ cells/T25 flask) were concultured at 37°C with syngeneic, irradiated (3,000 rads), peptide coated lymphoblasts ($2x10^6$ cells/well or $10x10^6$ cells/T25 flask) in R10 medium to give a final volume of 2 ml in 24-well plates or 10 ml in T25 flasks.

8. Restimulation of effector cells: Seven to ten days after the initial in vitro activation, described in paragraph 7 above, a portion of the effector cells were restimulated with irradiated (20,000 rads), peptide-coated Jurkat A2/K^b cells ($0.2x10^6$ cells/well) in the presence of $3x10^6$ "feeder cells"/well (C57Bl/6 irradiated spleen cells) in R10 medium supplemented with 5% rat ConA supernatant to help provide all of the cytokines needed for optimal effector cell growth.

9. Assay for cytotoxic activity: Target cells ($3x10^6$) were incubated at 37°C in the presence of 200 μl of sodium ^{51}Cr chromate. After 60 minutes, cells were washed three times and resuspended in R10 medium. Peptides were added at the required concentration. For the assay, 10^4 ^{51}Cr-labeled target cells wee added to different concentrations of effector cells (final volume of 200 μl) in U-bottom 96-2311 plates. After a 6-hour incubation period at 37°C, 0.1 ml aliquots of supernatant were removed from each well and radioactivity was determined in a Micromedic automatic gamma counter. The percent specific lysis was determined by the formula: percent specific release = 100x(experimental release - spontaneous release)/(maximum release - spontaneous release). Where peptide titrations wee performed, the antigenicity of a given peptide (for comparison purposes) was expressed as the peptide concentration required to induce 40% specific ^{51}Cr release at a given E:T.

Transgenic mice were injected subcutaneously in the base of the tail with an incomplete Freund's adjuvant emulsion containing 50 nM of the putative CTL epitopes containing the A2.1 motifs, and 50 nM of a hepatitis B core T helper epitope. Eight to 20 days later, animals were sacrificed and spleen cells were restimulated in vitro with syngeneic LPS lymphoblasts coated with the putative CTL epitope. A source of IL-2 (rat con A supernatant) was added at day 6 of the assay to a final concentration of 5% and CTL activity was measured on day 7. The capacity of these effector T cells to lyse peptide-coated target cells that express the A2 KB molecule (Jurkat A2 KB) was measured as lytic units. The results are presented in Table 24.

The results of this experiment indicate that those peptides having a binding of at least 0.01 are capable of inducing CTL. All of the peptides in Appendices 1 and 2 having a binding of at least about 0.01 would be immunogenic.

TABLE 24

Binding and Immunogenicity

HBV Polymerase (ayw)

Peptide	Binding**	CTL Activity	Algorithm
1 2 3 4 5 6 7 8 9			
F L L S L G I H L	0.52	63	-20.8
G L Y S S T V P V	0.15	10	-21.9
H L Y S H P I I L	0.13	10	-21.1
W I L R G T S F V	0.018	-+	-20.9
N L S W L S L D V	0.013	6	-24.7
L L S S N L S W L	0.005	-	-21.7
N L Q S L T N L L	0.003	-	-23.9
H L L V G S S G L	0.002	-	-24.7
L L D D E A G P L	0.0002	-	-25.5
P L E E E L P R L	0.0001	-	-26.1
D L N L G N L N V	-*	-	-25.7
N L Y V S L L L L	-	-	-23.6
P L P I H T A E L	-	-	-25.04

*-=<0.0001

** Relative binding capacity compared to std with IC_{50} = 52mM

xxx Lytic units/10^6 cells; 1 lytic unit = the number of effector cells required to give 30% Cr^{51} release.

-,-+ no measurable cytotoxic activity.

Example 11

Identification of immunogenic peptides

Using the motifs identified above for HLA-A2.1 allele amino acid sequences from a tumor-related protein, Melanoma Antigen-1 (MAGE-1), were analyzed for the presence of these motifs. Sequences for the target antigen are obtained from the GenBank data base (Release No. 71.0; 3/92). The identification of motifs is done using the "FINDPATTERNS" program (Devereux et al., Nucleic Acids Research 12:387-395 (1984)).

Other viral and tumor-related proteins can also be analyzed for the presence of these motifs. The amino acid sequence or the nucleotide sequence encoding products is obtained from the GenBank database in the cases of Human Papilloma Virus (HPV), Prostate Specific antigen (PSA), p53 oncogene, Epstein Barr Nuclear Antigen-1 (EBNA-1), and c-erb2 oncogene (also called HER-2/neu).

In the cases of Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), and Human Immunodeficiency Virus (HIV) several strains/isolates exist and many sequences have been placed in GenBank.

For HBV, binding motifs are identified for the adr, adw and ayw types. In order to avoid replication of identical sequences, all of the adr motifs and only those motifs from adw and ayw that are not present in adr are added to the list of peptides.

In the case of HCV, a consensus sequence from residue 1 to residue 782 is derived from 9 viral isolates. Motifs are identified on those regions that have no or very little (one residue) variation between the 9 isolates. The sequences of residues 783 to 3010 from 5 viral isolates were also analyzed. Motifs common to all the isolates are identified and added to the peptide list.

Finally, a consensus sequence for HIV type 1 for North American viral isolates (10-12 viruses) was obtained from the Los Alamos National Laboratory database (May 1991 release) and analyzed in order to identify motifs that are constant throughout most viral isolates. Motifs that bear a

small degree of variation (one residue, in 2 forms) were also added to the peptide list.

Appendices 1 and 2 provide the results of searches of the following antigens cERB2, EBNA1, HBV, HCV, HIV, HPV, MAGE, p53, and PSA. Only peptides with binding affinity of at least 1% as compared to the standard peptide in assays described in Example 5 are presented. Binding as compared to the standard peptide is shown in the far right column. The column labeled "Pos." indicates the position in the antigenic protein at which the sequence occurs.

Example 12

Identification of immunogenic peptides

Using the motifs disclosed here, amino acid sequences from various antigens were screened for further motifs. Screening was carried out as described in Example 11. Tables 25 and 26 provide the results of searches of the following antigens cERB2, CMV, Influenza A, HBV, HIV, HPV, MAGE, p53, PSA, Hu S3 ribosomal protein, LCMV, and PAP. Only peptides with binding affinity of at least 1% as compared to the standard peptide in assays described in Example 5 are presented. Binding as compared to the standard peptide is shown for each peptide.

TABLE 25

Sequence	Antigen	Molecule	A2 Bind.
KIFGSLAFL	c-ERB2		0.1500
RILHNGAYSL	c-ERB2		0.0180
IISAVVGILL	c-ERB2		0.0120
MMWFVVLTV	CMV		0.7600
YLLLYFSPV	CMV		0.7500
YLYRLNFCL	CMV		0.7200
FMWTYLVTL	CMV		0.6800
LLWWITILL	CMV		0.4900
GLWCVLFFV	CMV		0.4700
LMIRGVLEV	CMV		0.4500
LLLCRLPFL	CMV		0.4200
AMSRNLFRV	CMV		0.1500
AMLTACVEV	CMV		0.1000
RLQPNVPLV	CMV		0.0480
VLARTFTPV	CMV		0.0440
RLLRGLIRL	CMV		0.0370
WMWFPSVLL	CMV		0.0360
YLCCGITLL	CMV		0.0210
SLLTEVETYV	FLU-A	M1	0.0650
LLTEVETYV	FLU-A	M1	0.2000
LLTEVETYVL	FLU-A	M1	0.0130
GILGFVFTL	FLU-A	M1	0.1900
GILGFVFTLT	FLU-A	M1	0.0150
ILGFVFTLT	FLU-A	M1	0.2600
ILGFVFTLTV	FLU-A	M1	0.0550
ALASCMGLI	FLU-A	M1	0.0110
RMGAVTTEV	FLU-A	M1	0.0200

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
VTTEVAFGL	FLU-A	M1	0.0360
MVTTTNPLI	FLU-A	M1	0.0150
FTFSPTYKA	HBV	POL	0.0190
YLHTLWKAGI	HBV	POL	0.0280
LMLQAGFFLV	HBV(a)	ENV(a)	0.6300
RMLTIPQSV	HBV(a)	ENV(a)	0.0580
SLDSWWTSV	HBV(a)	ENV(a)	0.1000
FMLLLCLIFL	HBV(a)	ENV(a)	0.0450
LLPFVQWFV	HBV(a)	ENV(a)	0.6500
LMPFVQWFV	HBV(a)	ENV(a)	0.8300
FLGLSPTVWV	HBV(a)	ENV(a)	0.0300
SMLSPFLPLV	HBV(a)	ENV(a)	0.9700
GLWIRTPPV	HBV(a)	ENV(a)	0.3600
NLGNLNVSV	HBV(a)	ENV(a)	0.0160
YLHTLWKAGV	HBV(a)	POL(a)	0.1500
RLTGGVFLV	HBV(a)	POL(a)	0.1600
RMTGGVFLV	HBV(a)	POL(a)	0.1500
RLTGGVFLV	HBV(a)	ENV(a)	0.1600
ILGLLGFAV	HBV(a)	ENV(a)	0.0600
GLCQVFADV	HBV(a)	ENV(a)	0.0300
WLLRGTSFV	HBV(a)	ENV(a)	0.1000
YLPSALNPV	HBV(a)	ENV(a)	0.3200
LLVPFVQWFA	HBV adr		0.2600
FLPSDFFPSI	HBV adr		0.2100
VVSYVNVNM	HBV adr		0.0100
HLPDRVHFA	HBV adr		0.0160
SLAFSAVPA	HBV adr		0.0340

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
FLLTKILTI	HBV adw		0.6300
SLYNILSPFM	HBV adw		0.0440
CLFHIVNLI	HBV adw		0.2100
RLPDRVHFA	HBV adw		0.0940
ALPPASPSA	HBV adw		0.0710
GLLGWSPQA	HBV ayw		0.8650
FLGPLLVLQA	HBV ayw		0.0190
FLLTRILTI	HBV ayw		0.9300
GMLPVCPLI	HBV ayw		0.0520
QLFHLCLII	HBV ayw		0.0390
KLCLGWLWGM	HBV ayw		0.0210
LLWFHISCLI	HBV ayw		0.0130
YLVSFGVWI	HBV ayw		2.7000
LLEDWGPCA	HBV ayw		0.0180
KLHLYSHPI	HBV ayw		0.2900
FLLAQFTSA	HBV ayw		0.6600
LLAQFTSAI	HBV ayw		9.6000
YMDDVVLGA	HBV ayw		0.1600
ALMPLYACI	HBV ayw		0.2000
GLCQVFADA	HBV ayw		0.0180
HLPDLVHFA	HBV ayw		0.1100
RLCCQLDPA	HBV ayw		0.0290
ALMPLYACI	HBV ayw polymerase		0.5000
FLCKQYLNL	HBV ayw polymerase 665-673		0.0210

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
SLYADSPSV	HBV polymerase		0.3500
ALMPLYASI	HBV polymerase		0.0760
NLNNLNVSI	HBV polymerase		0.0660
ALSLIVNLL	HBV polymerase		0.0470
KLHLYSHPI	HBV polymerase		0.2900
WILRGTSFV	HBV polymerase 1344-1352		0.0270
LVLQAGFFLL	HBVadr	ENV	0.0150
FILLLCLIFL	HBVadr	ENV	0.0280
WILRGTSFV	HBVadr	POL	0.0180
IISCTCPTV	HBVadw	PreCore	0.0190
LVPFVQWFV	HBVadw	ENV	0.0200
LIISCSCPTV	HBVadw	CORE	0.0290
FLPSDFFPSI	HBVayr	PreCore	0.2100
LLCLGWLWGM	HBVayr	PreCore	0.0220
QLFHLCLII	HBVayw	PreCore	0.0390
CLGWLTGMDI	HBVayw	PreCore	0.0190
FLGGTTVCL	HBVayw	ENV	0.1700
SLYSILSPFL	HBVayw	ENV	0.2000
FLPSDFFPSV	HBVayw	CORE	1.5000
ILCWGELMTL	HBVayw	CORE	0.1900
LMTLATWVGV	HBVayw	CORE	0.6800
TLATWVGVNL	HBVayw	CORE	0.5700

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
GLSRYVARL	HBVayw	POL	0.1200
FLCKQYLNL	HBVayw	POL	0.1700
RMRGTFSAPL	HBVayw	POL	0.0110
SLYADSPSV	HBVayw	POL	0.3500
YLYGVGSAV	HCV		0.1600
LLSTTEWQV	HCV		0.0480
IIGAETFYV	HIV	POL	0.0260
QLWVTVYYGV	HIV	ENV	0.0250
NLWVTVYYGV	HIV	ENV	0.0160
KLWVTVYYGV	HIV	ENV	0.0150
KLWVTVYYGV	HIV.MN gp160		0.0150
YMLDLQPET	HPV16	E7	1.4000
TLGIVCPI	HPV16	E7	0.6500
YLLDLQEPV	HPV16(a)	E7 (a)	0.2200
YMLDLQPEV	HPV16(a)	E7 (a)	1.9000
MLDLQPETT	HPV16E7	E7	0.0130
SLQDIEITCVYCKTV	HPV18	E6	0.0100
RLLTSLFFL	HSV		0.3400
RLLTSLFFL	HSV		0.3400
LLLYYDYSL	HSV		0.2800
DMLGRVFFV	HSV		0.0110
TMFEALPHI	LCMV	Gp	0.2000
ALISFLLLA	LCMV	Gp	0.2200
TLMSIVSSL	LCMV	Gp	0.2000
NISGYNFSL	LCMV	Np	0.0280
ALLDGGNML	LCMV	Np	0.0320

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
ALHLFKTTV	LCMV	Gp	0.0170
SLISDQLLM	LCMV	Gp	0.0540
WLVTNGSYL	LCMV	Gp	0.0180
ALMDLLMFS	LCMV	Gp	0.4300
LMDLLMFST	LCMV	Gp	0.0460
LMFSTSAYL	LCMV	Gp	0.3600
YLVSIFLHL	LCMV	Gp	0.4200
SLHCKPEEA	MAGE1		0.0130
ALGLVCVQA	MAGE1		0.0150
LVLGTLEEV	MAGE1		0.0320
GTLEEVPTA	MAGE1		0.0130
CILESLFRA	MAGE1		0.0460
KVADLVGFLL	MAGE1		0.0560
KVADLVGFLLL	MAGE1		0.0200
VMIAMEGGHA	MAGE1		0.0360
SMHCKPEEV	MAGE1(a)		0.0180
AMGLVCVQV	MAGE1(a)		0.0120
LMLGTLEEV	MAGE1(a)		0.1300
KMADLVGFLV	MAGE1(a)		1.5000
VMVTCLGLSV	MAGE1(a)		0.3000
LLGDNQIMV	MAGE1(a)		0.0430
QMMPKTGFLV	MAGE1(a)		0.0500
VMIAMEGGHV	MAGE1(a)		0.0530
WMELSVMEV	MAGE1(a)		0.0410
FLWGPRALA	MAGE1N		0.0420
RALAETSYV	MAGE1N		0.0100
ALAETSYVKVL	MAGE1N		0.0120

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
ALAETSYVKV	MAGE1N		0.0150
KVLEYVIKV	MAGE1N		0.0900
YVIKVSARV	MAGE1N		0.0140
ALREEEEGV	MAGE1N		0.0210
YMFLWGPRV	MAGE1N(a)		0.2200
KMVELVHFLLL	MAGE2		0.6700
KMVELVHFL	MAGE2		0.1600
KMVELVHFLL	MAGE2		0.1100
KASEYLQLV	MAGE2		0.0110
YLQLVFGIEV	MAGE2		0.3700
LVFGIEVVEV	MAGE2		0.0120
QLVFGIELMEV	MAGE3		0.3400
KVAELVHFL	MAGE3		0.0550
KVAELVHFLL	MAGE3		0.0120
ELMEVDPIGHL	MAGE3		0.0260
HLYIFATCLGL	MAGE3		0.0410
IMPKAGLLIIV	MAGE3		0.0130
LVFGIELMEV	MAGE3		0.1100
ALGRNSFEV	p53 264-272 A8 (A1)		0.0570
LLGANSFEV	p53 264-272 A8 (A4)		0.1100
LLGRASFEV	p53 264-272 A8 (A5)		0.2200
LLGRNAFEV	p53 264-272 A8 (A6)		0.0390
LLGRNSFAV	p53 264-272 A8 (A8)		0.0420
RLGRNSFEV	p53 264-272 A8 (R1)		0.0190
LLGRRSFEV	p53 264-272 A8 (R5)		0.0540
LLGRNSFRV	p53 264-272 A8 (R8)		0.0250
LLFFWLDRSV	PAP		0.6000

Table 25 (Cont'd)

Sequence	Antigen	Molecule	A2 Bind.
VLAKELKFV	PAP		0.0590
ILLWQPIPV	PAP		1.3000
IMYSAHDTTV	PAP		0.0610
FLTLSVTWI	PSA		0.0150
FLTLSVTWIGA	PSA		0.0160
FLTLSVTWI	PSA		0.0150
VLVHPQWVLTA	PSA		0.0130
SLFHPEDTGQV	PSA		0.0190
MLLRLSEPAEL	PSA		0.1400
ALGTTCYA	PSA		0.0230
KLQCVDLHVI	PSA		0.0370
FLPSDYFPSV	HBVc18-27 analog		1.0000
YSFLPSDFFPSV	HBVc18-27 analog		0.0190

Table 26

Sequence	Antigen	Molecule	A2 Bind.
ALFLGFLGAA	HIV	gp160	0.4950
MLQLTVWGI	HIV	gp160	0.2450
RVIEVLQRA	HIV	gp160	0.1963
KLTPLCVTL	HIV	gp160	0.1600
LLIAARIVEL	HIV	gp160	0.1550
SLLNATDIAV	HIV	gp160	0.1050
ALFLGFLGA	HIV	gp160	0.0945
HMLQLTVWGI	HIV	gp160	0.0677
LLNATDIAV	HIV	gp160	0.0607
ALLYKLDIV	HIV	gp160	0.0362
WLWYIKIFI	HIV	gp160	0.0355
TIIVHLNESV	HIV	gp160	0.0350
LLQYWSQEL	HIV	gp160	0.0265
IMIVGGLVGL	HIV	gp160	0.0252
LLYKLDIVSI	HIV	gp160	0.0245
FLAIIWVDL	HIV	gp160	0.0233

Table 26 (Cont'd)

TLQCKIKQII	HIV	gp160	0.0200
GLVGLRIVFA	HIV	gp160	0.0195
FLGAAGSTM	HIV	gp160	0.0190
IISLWDQSL	HIV	gp160	0.0179
TVWGIKQLQA	HIV	gp160	0.0150
LLGRRGWEV	HIV	gp160	0.0142
AVLSIVNRV	HIV	gp160	0.0132
FIMIVGGLV	HIV	gp160	0.0131
LLNATDIAVA	HIV	gp160	0.0117
FLYGALLLA	PLP		1.9000
SLLTFMIAA	PLP		0.5300
FMIAATYNFAV	PLP		0.4950
RMYGVLPWI	PLP		0.1650
IAATYNFAV	PLP		0.0540
GLLECCARCLV	PLP		0.0515
YALTVVWLL	PLP		0.0415
ALTVVWLLV	PLP		0.0390
FLYGALLL	PLP		0.0345
SLCADARMYGV	PLP		0.0140
LLVFACSAV	PLP		0.0107

Table 26 (Cont'd)

Sequence	Antigen	A2
KMVELVHFLL	MAGE2	0.2200
KVAELVHFL	MAGE3	0.0550
RALAETSYV	MAGE1N	0.0100
LVFGIELMEV	MAGE3	0.1100
FLWGPRALA	MAGE1N	0.0420
ALAETSYVKV	MAGE1	0.0150
LVLGTLEEV	HIV	0.0320
LLWKGEGAVV	HIV	0.0360
IIGAETFYV	HIV	0.0260
LMVTVYYGV	HIV	0.4400
LLFNILGGWV	HCV	3.5000
LLALLSCLTV	HCV	0.6100
YLVAYQATV	HCV	0.2500
FLLLADARV	HCV	0.2300
ILAGYGAGV	HCV	0.2200
YLLPRRGPRL	HCV	0.0730
GLLGCIITSL	HCV	0.0610
DLMGYIPLV	HCV	0.0550
LLALLSCLTI	HCV	0.0340
VLAALAAYCL	HCV	0.0110
LLVPFVQWFV	HBV	1.6000
FLLAQFTSA	HBV	0.6600
FLLSLGIHL	HBV	0.5200
ALMPLYACI	HBV	0.5000
ILLLCLIFLL	HBV	0.3000
LLPIFFCLWV	HBV	0.1000
YLHTLWKAGI	HBV	0.0560

Table 26 (Cont'd)

YLHTLWKAGV	HBV	0.1300

Example 13

Identification of immunogenic peptides in autoantigens

As noted above, the motifs of the present invention can also be screened in antigens associated with autoimmune diseases. Using the motifs identified above for HLA-A2.1 allele amino acid sequences from myelin proteolipid (PLP), myelin basic protein (MBP), glutamic acid decarboxylase (GAD), and human collagen types II and IV were analyzed for the presence of these motifs. Sequences for the antigens were obtained from Trifilieff et al., *C.R. Sceances Acad. Sci.* 300:241 (1985); Eyler at al., *J. Biol. Chem.* 246:5770 (1971); Yamashita *et al. Biochiem. Biophys. Res. Comm.* 192:1347 (1993); Su *et al., Nucleic Acids Res.* 17:9473 (1989) and Pihlajaniemi *et al. Proc. Natl. Acad. Sci. USA* 84:940 (1987). The identification of motifs was done using the approach described in Example 5 and the algorithms of Examples 6 and 7. Table 27 provides the results of the search of these antigens.

Using the quantitative binding assays of Example 4, the peptides are next tested for the ability to bind MHC molecules. The ability of the peptides to suppress proliferative responses in autoreactive T cells is carried out using standard assays for T cell proliferation. For instance, methods as described by Miller *et al. Proc. Natl. Acad. Sci. USA*, 89:421 (1992) are suitable.

For further study, animal models of autoimmune disease can be used to demonstrate the efficacy of peptides of the invention. For instance, in HLA transgenic mice, autoimmune model diseases can be induced by injection of MBP, PLP or spinal cord homogenate (for MS), collagen (for arthritis). In addition, some mice become spontaneously affected by autoimmune disease (*e.g.*, NOD mice in diabetes). Peptides of the invention are injected into the appropriate animals, to identify preferred peptides.

TABLE 27

Human PLP peptides

Pos	AA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Allele	Motif
3	9	L	L	E	C	C	A	R	C	L		A2.1	(LM)2; (LVI)c
23	9	G	L	C	F	F	G	V	A	L			
39	9	A	L	T	G	T	E	K	L	I			
134	9	S	L	E	R	V	C	H	C	L			
145	9	W	L	G	H	P	D	K	F	V			
158	9	A	L	T	V	V	W	L	L	V			
164	9	L	L	V	F	A	C	S	A	V			
205	9	R	M	Y	G	V	L	P	W	I			
2	10	G	L	L	E	C	C	A	R	C	L		
3	10	L	L	E	C	C	A	R	C	L	V		
10	10	C	L	V	G	A	P	F	A	S	L		
163	10	W	L	L	V	F	A	C	S	A	V		
250	10	T	L	V	S	L	L	T	F	M	I		
64	9	V	I	H	A	F	Q	Y	V	I			Algorithm
80	9	F	L	Y	G	A	L	L	L	A			
157	9	Y	A	L	T	V	V	W	L	L			
163	9	W	L	L	V	F	A	C	S	A			
234	9	Q	M	T	F	H	L	F	I	A			
251	9	L	V	S	L	L	T	F	M	I			
253	9	S	L	L	T	F	M	I	A	A			
259	9	I	A	A	T	Y	N	F	A	V			
84	10	A	L	L	L	A	E	G	F	Y	T		
157	10	Y	A	L	T	V	V	W	L	L	V		
165	10	L	V	F	A	C	S	A	V	P	V		
218	10	K	V	C	G	S	N	L	L	S	I		
253	10	S	L	L	T	F	M	I	A	A	T		

Table 27 continued

Human Collagen TypeIV peptides

Pos	AA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Allele	Motif
5	9	A	L	M	G	P	L	G	L	L		A2.1	(LM)2; (LVI)c
11	9	G	L	L	G	Q	I	G	P	L			
23	9	G	M	L	G	Q	K	G	E	I			
231	9	P	L	G	Q	D	G	L	P	V			
3	10	T	L	A	L	M	G	P	L	G	L		
24	10	M	L	G	Q	K	G	E	I	G	L		
59	10	P	L	G	K	D	G	P	P	G	V		
139	10	P	L	G	L	P	G	A	S	G	L		

Human Collagen TypeII peptides

Pos	AA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Allele	Motif
794	9	G	L	A	G	Q	R	G	I	V		A2.1	(LM)2; (LVI)c
17	9	V	M	Q	G	P	M	G	P	M			Algorithm

Table 27 continued

Human GAD peptides

Pos	AA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Allele	Motif
56	9	S	L	E	E	K	S	R	L	V		A2.1	(LM)2; (LVI)c
116	9	F	L	L	E	V	V	D	I	L			
117	9	L	L	E	V	V	D	I	L	L			
150	9	G	M	E	G	F	N	L	E	L			
157	9	E	L	S	D	H	P	E	S	L			
168	9	I	L	V	D	C	R	D	T	L			
190	9	Q	L	S	T	G	L	D	I	I			
229	9	T	L	K	K	M	R	E	I	V			
275	9	G	M	A	A	V	P	K	L	V			
300	9	A	L	G	F	G	T	D	N	V		A2.1	(LM)2; (LVI)c
409	9	V	L	L	Q	C	S	A	I	L			
410	9	L	L	Q	C	S	A	I	L	V			
416	9	I	L	V	K	E	K	G	I	L			
466	9	L	M	W	K	A	K	G	T	V			
534	9	K	L	H	K	V	A	P	K	I			
546	9	M	M	E	S	G	T	T	M	V			
582	9	F	L	I	E	E	I	E	R	L			
42	10	K	L	G	L	K	I	C	G	F	L		
116	10	F	L	L	E	V	V	D	I	L	L		
138	10	V	L	D	F	H	H	P	H	Q	L		
147	10	L	L	E	G	M	E	G	F	N	L		
212	10	N	M	F	T	Y	E	I	A	P	V		
275	10	G	M	A	A	V	P	K	L	V	L		
300	10	A	L	G	F	G	T	D	N	V	I		
328	10	I	L	E	A	K	Q	K	G	Y	V		
381	10	L	M	S	R	K	H	R	H	K	L		
409	10	V	L	L	Q	C	S	A	I	L	V		
435	10	L	L	Q	P	D	K	Q	Y	D	V		
465	10	W	L	M	W	K	A	K	G	T	V		
485	10	E	L	A	E	Y	L	Y	A	K	I		
545	10	L	M	M	E	S	G	T	T	M	V		
252	9	G	A	I	S	N	M	Y	S	I			Algorithm
367	9	N	L	W	L	H	V	D	A	A			
567	9	R	M	V	I	S	N	P	A	A			
299	10	A	A	L	G	F	G	T	D	N	V		
406	10	M	M	G	V	L	L	Q	C	S	A		
423	10	I	L	Q	G	C	N	Q	M	C	A		

Example 14

Immunogenicity of HPV peptides in A2.1 transgenic mice

A group of 14 HPV peptides, including 9 potential epitopes plus 3 low binding and one non-binding peptides as controls was screened for immunogenicity in HLA-A2.1 transgenic mice using the methods described in Example 10. To test the immunogenic potential of the peptides, HLA A2.1 transgenic mice were injected with 50 μg/mouse of each HPV peptide together with 140 μg/mouse of helper peptide (HBV core 128-140 (TPPAYRPPNAPIL). The peptides were injected in the base of the tail in a 1:1 emulsion IFA. Three mice per group were used. As a positive control, the HBV polymerase 561-570 peptide, which induced a strong CTL response in previous experiments, was utilized.

Based on these results (Table 28), four unrelated peptides were considered to be the most immunogenic: TLGIVCPI , LLMGTLGIV, YMLDLQPETT, and TIHDIILECV. TLGIVCPI and YMLDLQPETT were found to be good HLA-A2.1 binders, while LLMGTLGIV and TIHDIILECV were found to be intermediate binders in previous binding assays.

TABLE 28

HPV-16 Peptides for possible use in clinical trial

Peptide Position/ Cytel ID	Sequence	AA	A2.1 binding	Immunogenicity Experiment 1	Immunogenicity Experiment 2
E7.86/1088.01	TLGIVCPI	8	0.15	94.4 (1.34)	54.2 (1.43)*
E7.86/1088.06	TLGIVCPIC	9	0.075	2.05 (4.93)	1.3 (3.74)
E7.85/1088.08	GTLGIVCPI	9	0.021	9/08 (3.93)	-**
E7.11/1088.03	YMLDLQPETT	10	0.15	10.32 (1.66)	5.7 (2.39)
E7.11/1088.04	YMLDLQPET	9	0.14	5.0 (3.70)	2.6 (15.5)
E7.12/1088.09	MLDLQPETT	9	0.0028	-	-
E6.52/1088.05	FAFRDLCIV	9	0.057	-	ND
E7.82/1088.02	LLMGTLGIV	9	0.024	9.62 (2.53)	8.93 (1.91)
E6.29/1088.10	TIHDIILECV	10	0.021	22.13 (3.71)	0.4 (3.52)
E7.7/1088.07	TLHEYMLDL	9	0.0070	-	1.2 (3.88
E6.18/1088.15	KLPQLCTEL	9	0.0009	-	0.3 (5.64)
E6.7/1088.11	AMFQDPQER	10	0.0002	-	ND
E6.26/1088.12	LQTTIHDII	9	0.0002	-	-
E7.73/1088.13	HVDIRTLED	9	0	-	ND

* Δ Lytic Units, geometric mean x+ SD (3 mice/peptide)
** a dash indicates Δ Lytic Units with a geometric mean ≤0.2

Mixtures of selected HPV epitopes

A combination of CTL peptides and a helper peptide were tested for the ability to provide an increased immune response. The four single peptides were injected separately in order to compare their immunogenicity to injections containing only the two good binders or only the two intermediate binders. In addition all four peptide were injected together. To further evaluate the immunogenicity of a combination of peptides with different binding affinity decreases, another control was introduced in this experiment. A mixture of the two good binders was injected in a different site than the mixture of the two intermediate binders into the base of the tail of the same mouse. All groups of CTL epitopes were injected together with the HBVc helper epitope, with the exception of two groups in which all four HPV coinjected with two different doses of a PADRE helper peptide (aKXVAAWTLKAAa, where a is d-alanine and X is cyclohexylalanine) either 1μg or 0.05μg per mouse.

All four peptides induced a strong CTL response when injected alone and tested using target cells labeled with the appropriate peptide (Table 29). TLGIVCPI proved to be the strongest epitope, an observation confirming the results described above. When mixtures of all four peptides were injected and the responses were stimulated in vitro and tested with target cells pulsed with each single peptide, all combinations showed a strong CTL response. No significant difference was observed when the two helper epitopes were compared. This might in part be due to the fact that the highest dose of PADRE used in this experiment was 140-fold lower than the one for the HBV helper peptide.

Injection of mixtures of the two good binders together or the two intermediate binders resulted in a very low CTL response in both cases even though the single peptides were highly effective. These results, however, are due to a very low number of cell recovery after splenocyte culture of 6 days and are therefore regarded as preliminary.

TABLE 29

HPV Peptides single and in combinations

A

Peptide/s injected	Peptides in restimulation and CTL assay 1088.01	1088.02	1088.03	1088.10
same as in vitro	116.1 (3.49)*	55.98 (2.49)	5.56 (1.75)	16.4 (1.49)
1088.01 + 1088.03 + 875.23	1.37 (16.56)		0 (0)	
1088.02 + 1088.10 + 875.23		1.11 (2.9)		1.62 (13.1)
1088.01/.03 + 1088.02/.10 + 875.23	19.5 (4.1)	4.68 (2.3)	1.13 (21.9)	1.17 (2.58)
1088.all + 875.23	107.9 (4.77)	13.52 (1.4)	2.58 (5.07)	102.3 (1.32)
1088.all + PADRE 1 μg	73.11 (4.48)	16.83 (2.54)	3.55 (2.9)	20.13 (1.05)
1088.all + PADRE 0.05 μg	37.15 (2.25)	26.79(2.09)	6.5 (1.64)	4.45 (4.14)

* Δ Lytic Units 30% geometric mean (+x deviation)

Peptides were dissolved in 50%DMSO/H2O to reach a stock concentration of 20mg/ml and were further dissolved in sterile PBS. For subcutaneous injection in the base of the tail of A2.1 transgenic mice, the peptide solution was mixed 1:1 with IFA. The injected amount of HPV-CTL peptides was 50 μg/mouse coinjected with 140 μg/mouse of the HBVcore peptide 875.23 or the indicated dose of PADRE (3 mice/group). Spleens were removed on day 11 and splenocytes were restimulated in vitro with irradiated LPS-Blasts pulsed with the indicated HPV-CTL epitopes at 1μg/ml. After six days, the cytotoxic assay was performed using Jurkat JA2Kb cells (A) or MBB17 (B) as target cells labelled with 51Cr in the presence or absence of the appropriate HPV epitope peptides.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0841	ILSPFLPLL	9	HBV	adr	ENV	371	2.9
1.0240	TLQDIVLHL	9	HPV	18	E7	7	0.76
1.0838	WLSLLVPFV	9	HBV	adr	ENV	335	0.72
1.0851	FLLSLGIHL	9	HBV	adr	POL	1147	0.52
1.0306	QLFEDNYAL	9	c-ERB2			106	0.46
1.0814	LMVTVYYGV	9	HIV		ENV	2182	0.44
1.0878	MMWFWGPSL	9	HBV	adw	ENV	360	0.41
1.0839	MMWYWGPSL	9	HBV	adr	ENV	360	0.41
1.0384	FLTKQYLNL	9	HBV	adw	POL	1279	0.29
1.0321	ILHNGAYSL	9	c-ERB2			435	0.21
1.0834	LLLCLIFLL	9	HBV	adr	ENV	250	0.19
1.0167	GLYSSTVPV	9	HBV	adr	POL	635	0.15
1.0849	HLYSHPIIL	9	HBV	adr	POL	1076	0.13
1.0275	RMPEAAPPV	9	p53			65	0.12
1.0854	LLMGTLGIV	9	HPV	16	E7	82	0.11
1.0880	ILSPFMPLL	9	HBV	adw	ENV	371	0.11
1.0127	YLVAYQATV	9	HCV		LORF	1585	0.11
1.0151	VLLDYQGML	9	HBV	adr	ENV	259	0.11
1.0018	VLAEAMSQV	9	HIV		GAG	367	0.11
1.0330	RLLQETELV	9	c-ERB2			689	0.091
1.0209	SLYAVSPSV	9	HBV	adr	POL	1388	0.078
1.0816	DLMGYIPLV	9	HCV		CORE	132	0.055
1.0835	LLCLIFLLV	9	HBV	adr	ENV	251	0.049
1.0852	FLCQQYLHL	9	HBV	adr	POL	1250	0.048

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0882	NLYVSLMLL	9	HBV	adw	POL	1088	0.046
1.0837	GMLPVCPLL	9	HBV	adr	ENV	265	0.046
1.0819	ILPCSFTTL	9	HCV		NS1/ENV2	676	0.045
1.0109	ALSTGLIHL	9	HCV		NS1/ENV2	686	0.042
1.0833	ILLLCLIFL	9	HBV	adr	ENV	249	0.035
1.0301	HLYQGCQVV	9	c-ERB2			48	0.034
1.0337	CLTSTVQLV	9	c-ERB2			789	0.034
1.0842	PLLPIFFCL	9	HBV	adr	ENV	377	0.031
1.0861	ALCRWGLLL	9	c-ERB2			5	0.031
1.0309	VLIQRNPQL	9	c-ERB2			153	0.029
1.0828	VLQAGFFLL	9	HBV	adr	ENV	177	0.024
1.0844	LLWFHISCL	9	HBV	adr	CORE	490	0.024
1.0135	ILAGYGAGV	9	HCV		LORF	1851	0.024
1.0870	QLMPYGCLL	9	c-ERB2			799	0.023
1.0075	LLWKGEGAV	9	HIV		POL	1496	0.023
1.0873	FLGGTPVCL	9	HBV	adw	ENV	204	0.021
1.0323	ALIHHNTHL	9	c-ERB2			466	0.021
1.0859	VLVHPQWVL	9	PSA			49	0.020
1.0267	KLQCVDLHV	9	PSA			166	0.019
1.0820	VLPCSFTTL	9	HCV		NS1/ENV2	676	0.017
1.0111	HLHQNIVDV	9	HCV		NS1/ENV2	693	0.016
1.0103	SMVGNWAKV	9	HCV		ENV1	364	0.016
1.0283	LLGRNSFEV	9	p53			264	0.014
1.0207	GLYRPLLSL	9	HBV	adr	POL	1370	0.014

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0389	GLYRPLLRL	9	HBV	adw	POL	1399	0.014
1.0185	NLSWLSLDV	9	HBV	adr	POL	996	0.013
1.0113	FLLLADARV	9	HCV		NS1/ENV2	725	0.013
1.0119	YLVTRHADV	9	HCV		LORF	1131	0.011
1.0846	CLTHIVNLL	9	HBV	adr	POL	912	0.010
1.0156	ELMNLATWV	9	HBV	adr	CORE	454	0.010
1.0236	KLPDLCTEL	9	HPV	18	E6	13	0.010
1.0056	ALQDSGLEV	9	HIV		POL	1180	0.0083
1.0375	LLSSDLSWL	9	HBV	adw	POL	1021	0.0081
1.0094	ALAHGVRVL	9	HCV		CORE	150	0.0072
1.0129	TLHGPTPLL	9	HCV		LORF	1617	0.0070
1.0041	KLLRGTKAL	9	HIV		POL	976	0.0069
1.0131	CMSADLEVV	9	HCV		LORF	1648	0.0067
1.0872	GLLGPLLVL	9	HBV	adw	ENV	170	0.0066
1.0228	TLHEYMLDL	9	HPV	16	E7	7	0.0059
1.0274	KLLPENNVL	9	p53			24	0.0058
1.0043	ILKEPVHGV	9	HIV		POL	1004	0.0055
1.0206	RLGLYRPLL	9	HBV	adr	POL	1368	0.0050
1.0188	GLPRYVARL	9	HBV	adr	POL	1027	0.0050
1.0202	KLIGTDNSV	9	HBV	adr	POL	1317	0.0050
1.0818	FLLALLSCL	9	HCV		CORE	177	0.0046
1.0184	LLSSNLSWL	9	HBV	adr	POL	992	0.0046
1.0102	QLLRIPQAV	9	HCV		ENV1	337	0.0039
1.0114	GLRDLAVAV	9	HCV		LORF	963	0.0034

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0005	TLNAWVKVI	9	HIV		GAG	156	0.0032
1.0183	NLQSLTNLL	9	HBV	adr	POL	985	0.0025
1.0359	QLGRKPTPL	9	HBV	adw	ENV	89	0.0025
1.0150	SLDSWWTSL	9	HBV	adr	ENV	194	0.0023
1.0362	ILSKTGDPV	9	HBV	adw	ENV	153	0.0021
1.0866	ILLVVVLGV	9	c-ERB2			661	0.0020
1.0214	LLHKRTLGL	9	HBV	adr	"X"	1510	0.0019
1.0216	CLFKDWEEL	9	HBV	adr	"X"	1533	0.0019
1.0862	GLGISWLGL	9	c-ERB2			447	0.0018
1.0187	HLLVGSSGL	9	HBV	adr	POL	1020	0.0018
1.0318	TLEEITGYL	9	c-ERB2			402	0.0018
1.0328	PLTSIISAV	9	c-ERB2			650	0.0015
1.0822	LLGCIITSL	9	HCV		LORF	1039	0.0015
1.0277	ALNKMFCQL	9	p53			129	0.0013
1.0066	HLEGKIILV	9	HIV		POL	1322	0.0010
1.0308	QLRSLTEIL	9	c-ERB2			141	0.0008
1.0115	DLAVAVEPV	9	HCV		LORF	966	0.0008
1.0391	VLHKRTLGL	9	HBV	adw	"X"	1539	0.0007
1.0876	FLCILLLCL	9	HBV	adw	ENV	246	0.0007
1.0148	LLDPRVRGL	9	HBV	adr	ENV	120	0.0006
1.0221	KLPQLCTEL	9	HPV	16	E6	18	0.0006
1.0065	HLEGKVILV	9	HIV		POL	1322	0.0006
1.0017	EMMTACQGV	9	HIV		GAG	350	0.0006
1.0055	HLALQDSGL	9	HIV		POL	1178	0.0005

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0868	VLGVVFGIL	9	c-ERB2			666	0.0005
1.0004	TLNAWVKVV	9	HIV		GAG	156	0.0005
1.0381	HLESLYAAV	9	HBV	adw	POL	1165	0.0005
1.0128	CLIRLKPTL	9	HCV		LORF	1610	0.0004
1.0255	CLGLSYDGL	9	MAGE	1/3		174	0.0004
1.0212	HLSLRGLPV	9	HBV	adr	"X"	1470	0.0004
1.0247	ILESLFRAV	9	MAGE	1		93	0.0004
1.0092	TLTCGFADL	9	HCV		CORE	125	0.0003
1.0108	TLPALSTGL	9	HCV		NS1/ENV2	683	0.0003
1.0294	ALAIPQCRL	9	EBNA1			525	0.0003
1.0101	DLCGSVFLV	9	HCV		ENV1	280	0.0003
1.0231	RLCVQSTHV	9	HPV	16	E7	66	0.0003
1.0162	LLDDEAGPL	9	HBV	adr	POL	587	0.0002
1.0829	CLRRFIIFL	9	HBV	adr	ENV	239	0.0002
1.0126	GLPVCQDHL	9	HCV		LORF	1547	0.0001
1.0163	PLEEELPRL	9	HBV	adr	POL	594	0.0001
1.0130	PLLYRLGAV	9	HCV		LORF	1623	0.0001
1.0042	ELAENREIL	9	HIV		POL	997	0
1.0054	ELQAIHLAL	9	HIV		POL	1173	0
1.0089	LIPRRGPRL	9	HCV		CORE	36	0
1.0091	NLGKVIDTL	9	HCV		CORE	118	0
1.0093	PLGGAARAL	9	HCV		CORE	143	0
1.0154	DLLDTASAL	9	HBV	adr	CORE	419	0
1.0178	QLKQSRLGL	9	HBV	adr	POL	791	0

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0179	GLQPQQGSL	9	HBV	adr	POL	798	0
1.0286	PLDGEYFTL	9	p53			322	0
1.0296	VLKDAIKDL	9	EBNA1			574	0
1.0310	QLCYQDTIL	9	c-ERB2			160	0
1.0007	DLNTMLNTV	9	HIV		GAG	188	0
1.0037	ELHPDKWTV	9	HIV		POL	928	0
1.0070	ELKKIIGQV	9	HIV		POL	1412	0
1.0157	ELVVSYVNV	9	HBV	adr	CORE	473	0
1.0160	CLTFGRETV	9	HBV	adr	CORE	497	0
1.0164	DLNLGNLNV	9	HBV	adr	POL	614	0
1.0867	LLVVVLGVV	9	c-ERB2			662	0
1.0159	NMGLKIRQL	9	HBV	adr	CORE	482	0
1.0322	SLRELGSGL	9	c-ERB2			457	<0.0002
1.0350	DLLEKGERL	9	c-ERB2			933	<0.0002
1.0352	DLVDAEEYL	9	c-ERB2			1016	<0.0002
1.0366	PLEEELPHL	9	HBV	adw	POL	623	<0.0002
1.0372	DLQHGRLVL	9	HBV	adw	POL	781	<0.0002
1.0390	PLPGPLGAL	9	HBV	adw	"X"	1476	<0.0002
1.0811	LLTQIGCTL	9	HIV		POL	685	<0.0002
1.0812	PLVKLWYQL	9	HIV		POL	1116	<0.0002
1.0832	FLFILLLCL	9	HBV	adr	ENV	246	<0.0002
1.0847	NLYVSLLLL	9	HBV	adr	POL	1059	<0.0002
1.0316	PLQPEQLQV	9	c-ERB2			391	<0.0002
1.0342	DLAARNVLV	9	c-ERB2			845	<0.0002

APPENDIX I: 9-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0343	VLVKSPNHV	9	c-ERB2			851	<0.0002
1.0356	TLSPGKNGV	9	c-ERB2			1172	<0.0002
1.0376	DLSWLSLDV	9	HBV	adw	POL	1025	<0.0002
1.0363	NMENIASGL	9	HBV	adw	ENV	163	<0.0002
1.0195	TLPQEHIVL	9	HBV	adr	POL	1179	<0.0003
1.0196	KLKQCFRKL	9	HBV	adr	POL	1188	<0.0003
1.0201	PLPIHTAEL	9	HBV	adr	POL	1296	<0.0003
1.0210	QLDPARDVL	9	HBV	adr	"X"	1426	<0.0003
1.0220	VLGGCRHKL	9	HBV	adr	"X"	1551	<0.0003
1.0229	DLQPETTDL	9	HPV	16	E7	14	<0.0003
1.0245	ALEAQQEAL	9	MAGE	1		15	<0.0003
1.0266	DLPTQEPAL	9	PSA			136	<0.0003
1.0279	HLIRVEGNL	9	p53			193	<0.0003
1.0282	TLEDSSGNL	9	p53			256	<0.0003
1.0238	ELRHYSDSV	9	HPV	18	E6	77	<0.0003
1.0268	DLHVISNDV	9	PSA			171	<0.0003
1.0836	CLIFLLVLL	9	HBV	adr	ENV	253	<0.0006

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0890	LLFNILGGWV	10	HCV		LORF	1807	3.5
1.0930	LLVPFVQWFV	10	HBV	adw	ENV	338	1.6
1.0884	LLALLSCLTV	10	HCV		CORE	178	0.61
1.0896	ILLLCLIFLL	10	HBV	adr	ENV	249	0.30
1.0518	GLSPTVWLSV	10	HBV	adr	ENV	348	0.28
1.0902	SLYNILSPFL	10	HBV	adr	ENV	367	0.23
1.0892	LLVLQAGFFL	10	HBV	adr	ENV	175	0.21
1.0686	FLQTHIFAEV	10	EBNA1			565	0.17
1.0628	QLFLNTLSFV	10	HPV	18	E7	88	0.11
1.0904	LLPIFFCLWV	10	HBV	adr	ENV	378	0.10
1.0897	LLLCLIFLLV	10	HBV	adr	ENV	250	0.099
1.0516	LLDYQGMLPV	10	HBV	adr	ENV	260	0.085
1.0901	WMMWYWGPSL	10	HBV	adr	ENV	359	0.084
1.0533	GLYSSTVPVL	10	HBV	adr	POL	635	0.080
1.0469	YLLPRRGPRL	10	HCV		CORE	35	0.073
1.0888	GLLGCIITSL	10	HCV		LORF	1038	0.061
1.0907	ILCWGELMNL	10	HBV	adr	CORE	449	0.052
1.0927	LLGICLTSTV	10	c-ERB2			785	0.049
1.0452	LLWKGEGAVV	10	HIV		POL	1496	0.036
1.0885	LLALLSCLTI	10	HCV		CORE	178	0.034
1.0620	KLTNTGLYNL	10	HPV	18	E6	92	0.032
1.0502	RLIVFPDLGV	10	HCV		LORF	2578	0.032
1.0659	FLTPKKLQCV	10	PSA			161	0.031
1.0932	WMMWFWGPSL	10	HBV	adw	ENV	359	0.029

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0772	SLNFLGGTPV	10	HBV	adw	ENV	201	0.027
1.0609	SLQDIEITCV	10	HPV	18	E6	24	0.025
1.0526	ILSTLPETTV	10	HBV	adr	CORE	529	0.022
1.0508	RLHGLSAFSL	10	HCV		LORF	2885	0.020
1.0493	ILGGWVAAQL	10	HCV		LORF	1811	0.018
1.0738	VMAGVGSPYV	10	c-ERB2			773	0.018
1.0460	QLMVTVYYGV	10	HIV		ENV	2181	0.017
1.0573	ILRGTSFVYV	10	HBV	adr	POL	1345	0.016
1.0703	SLTEILKGGV	10	c-ERB2			144	0.015
1.0912	LLGCAANWIL	10	HBV	adr	POL	1337	0.014
1.0798	ALPPASPSAV	10	HBV	adw	"X"	1483	0.013
1.0908	QLLWFHISCL	10	HBV	adr	CORE	489	0.013
1.0677	NLLGRNSFEV	10	p53			263	0.013
1.0889	VLAALAAYCL	10	HCV		LORF	1666	0.011
1.0528	LLLDDEAGPL	10	HBV	adr	POL	586	0.011
1.0500	IMAKNEVFCV	10	HCV		LORF	2558	0.0088
1.0492	VLVGGVLAAL	10	HCV		LORF	1661	0.0084
1.0898	LLCLIFLLVL	10	HBV	adr	ENV	251	0.0075
1.0458	KLMVTVYYGV	10	HIV		ENV	2181	0.0069
1.0459	NLMVTVYYGV	10	HIV		ENV	2181	0.0067
1.0530	GLSPTVWLSA	10	HBV	adw	ENV	348	0.0067
1.0759	SLPTHDPSPL	10	c-ERB2			1100	0.0059
1.0419	VLPEKDSWTV	10	HIV		POL	940	0.0056
1.0666	FLHSGTAKSV	10	p53			113	0.0050

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0473	GLIHLHQNIV	10	HCV		NS1/ENV2	690	0.0047
1.0792	SLYAAVTNFL	10	HBV	adw	POL	1168	0.0046
1.0780	IMPARFYPNV	10	HBV	adw	POL	713	0.0043
1.0507	YLTRDPTTPL	10	HCV		LORF	2803	0.0042
1.0914	GLYNLLIRCL	10	HPV	18	E6	97	0.0036
1.0649	YLEYGRCRTV	10	MAGE	1		248	0.0034
1.0561	SLFTSITNFL	10	HBV	adr	POL	1139	0.0034
1.0788	NLLSSDLSWL	10	HBV	adw	POL	1020	0.0032
1.0753	RMARDPQRFV	10	c-ERB2			978	0.0020
1.0568	RMRGTFVVPL	10	HBV	adr	POL	1288	0.0020
1.0642	SLQLVFGIDV	10	MAGE	1		150	0.0020
1.0582	KLLHKRTLGL	10	HBV	adr	'X'	1509	0.0019
1.0713	GLGMEHLREV	10	c-ERB2			344	0.0017
1.0742	GMSYLEDVRL	10	c-ERB2			832	0.0017
1.0549	NLLSSNLSWL	10	HBV	adr	POL	991	0.0016
1.0465	QLTVWGIKQL	10	HIV		ENV	2760	0.0015
1.0524	VLEYLVSFGV	10	HBV	adr	CORE	505	0.0015
1.0483	VLNPSVAATL	10	HCV		LORF	1253	0.0015
1.0548	SLTNLLSSNL	10	HBV	adr	POL	988	0.0014
1.0512	ALLDPRVRGL	10	HBV	adr	ENV	119	0.0011
1.0676	TLEDSSGNLL	10	p53			256	0.0011
1.0719	TLQGLGISWL	10	c-ERB2			444	0.0011
1.0627	DLRAFQQLFL	10	HPV	18	E7	82	0.0010
1.0725	VLQGLPREYV	10	c-ERB2			546	0.0009

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0918	DLPPWFPPMV	10	EBNA1			605	0.0009
1.0499	DLSDGSWSTV	10	HCV		LORF	2399	0.0008
1.0559	CLAFSYMDDV	10	HBV	adr	POL	1118	0.0008
1.0632	PLVLGTLEEV	10	MAGE	1		37	0.0008
1.0520	NLATWVGSNL	10	HBV	adr	CORE	457	0.0008
1.0400	NLLTQIGCTL	10	HIV		POL	684	0.0007
1.0488	GLTHIDAHFL	10	HCV		LORF	1564	0.0007
1.0733	VLGSGAFGTV	10	c-ERB2			725	0.0007
1.0434	QLIKKEKVYL	10	HIV		POL	1219	0.0006
1.0451	KLLWKGEGAV	10	HIV		POL	1495	0.0006
1.0470	SMVGNWAKVL	10	HCV		ENV1	364	0.0006
1.0570	KLIGTDNSVV	10	HBV	adr	POL	1317	0.0006
1.0924	ILLVVVLGVV	10	c-ERB2			661	0.0006
1.0397	LLDTGADDTV	10	HIV		POL	619	0.0005
1.0446	HLKTAVQMAV	10	HIV		POL	1426	0.0005
1.0604	DLLMGTLGIV	10	HPV	16	E7	81	0.0005
1.0443	LLKLAGRWPV	10	HIV		POL	1356	0.0004
1.0461	DLMVTVYYGV	10	HIV		ENV	2181	0.0004
1.0619	TLEKLTNTGL	10	HPV	18	E6	89	0.0004
1.0787	SLTNLLSSDL	10	HBV	adw	POL	1017	0.0004
1.0521	NLEDPASREL	10	HBV	adr	CORE	465	0.0003
1.0583	GLSAMSTTDL	10	HBV	adr	'X'	1517	0.0003
1.0652	VLVASRGRAV	10	PSA			36	0.0003
1.0716	DLSVFQNLQV	10	c-ERB2			421	0.0003

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos	A2.1
1.0723	QLFRNPHQAL	10	c-ERB2			484	0.0003
1.0727	PLTSIISAVV	10	c-ERB2			650	0.0003
1.0479	YLKGSSGGPL	10	HCV		LORF	1160	0.0002
1.0497	QLPCEPEPDV	10	HCV		LORF	2159	0.0002
1.0523	CLTFGRETVL	10	HBV	adr	CORE	497	0.0002
1.0603	TLEDLLMGTL	10	HPV	16	E7	78	0.0002
1.0631	SLHCKPEEAL	10	MAGE	1		7	0.0002
1.0680	EMFRELNEAL	10	p53			339	0.0002
1.0689	VLKDAIKDLV	10	EBNA1			574	0.0002
1.0757	DLVDAEEYLV	10	c-ERB2			1016	0.0002
1.0796	RMRGTFVSPL	10	HBV	adw	POL	1317	0.0002
1.0669	QLAKTCPVQL	10	p53			136	0.0001
1.0717	NLQVIRGRIL	10	c-ERB2			427	0.0001
1.0721	WLGLRSLREL	10	c-ERB2			452	0.0001
1.0522	NMGLKIRQLL	10	HBV	adr	CORE	482	0
1.0527	PLSYQHFRKL	10	HBV	adr	POL	576	0
1.0529	ELPRLADEGL	10	HBV	adr	POL	598	0
1.0531	GLNRRVAEDL	10	HBV	adr	POL	606	0
1.0536	PLTVNEKRRL	10	HBV	adr	POL	672	0
1.0539	IMPARFYPNL	10	HBV	adr	POL	684	0
1.0550	PLHPAAMPHL	10	HBV	adr	POL	1012	0
1.0552	DLHDSCSRNL	10	HBV	adr	POL	1051	0
1.0555	LLYKTFGRKL	10	HBV	adr	POL	1066	0
1.0557	PMGVGLSPFL	10	HBV	adr	POL	1090	0

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0560	VLGAKSVQHL	10	HBV	adr	POL	1128	0
1.0569	PLPIHTAELL	10	HBV	adr	POL	1296	0
1.0579	PLPSLAFSAV	10	HBV	adr	'X'	1454	0
1.0585	DLEAYFKDCL	10	HBV	adr	'X'	1525	0
1.0587	ELGEEIRLKV	10	HBV	adr	'X'	1540	0
1.0589	VLGGCRHKLV	10	HBV	adr	'X'	1551	0
1.0597	TLEQQYNKPL	10	HPV	16	E6	94	0
1.0608	DLCTELNTSL	10	HPV	18	E6	16	0
1.0616	RLQRRRETQV	10	HPV	18	E6	49	0
1.0621	HLEPQNEIPV	10	HPV	18	E7	14	0
1.0639	LLKYRAREPV	10	MAGE	1/3		114	0
1.0643	CLGLSYDGLL	10	MAGE	1/3		174	0
1.0657	DMSLLKNRFL	10	PSA			98	0
1.0658	LLRLSEPAEL	10	PSA			119	0
1.0663	PLSQETFSDL	10	p53			13	0
1.0664	PLPSQAMDDL	10	p53			34	0
1.0690	ELAALCRWGL	10	c-ERB2			2	0
1.0692	RLPASPETHL	10	c-ERB2			34	0
1.0699	RLRIVRGTQL	10	c-ERB2			98	0
1.0701	GLRELQLRSL	10	c-ERB2			136	0
1.0730	QMRILKETEL	10	c-ERB2			711	0
1.0732	ILKETELRKV	10	c-ERB2			714	0
1.0754	PLDSTFYRSL	10	c-ERB2			999	0
1.0755	LLEDDDMGDL	10	c-ERB2			1008	0

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0758	DLGMGAAKGL	10	c-ERB2			1089	9
1.0761	PLPSETDGYV	10	c-ERB2			1119	0
1.0763	TLSPGKNGVV	10	c-ERB2			1172	0
1.0765	TLQDPRVRAL	10	HBV	adw	ENV	119	0
1.0768	NMENIASGLL	10	HBV	adw	ENV	163	0
1.0775	ELPHLADEGL	10	HBV	adw	POL	627	0
1.0776	GLNRPVAEDL	10	HBV	adw	POL	635	0
1.0777	PLTVNENRRL	10	HBV	adw	POL	701	0
1.0790	LLYKTYGRKL	10	HBV	adw	POL	1095	0
1.0801	GLSAMSPTDL	10	HBV	adw	"X"	1546	0
1.0802	DLEAYFKDCV	10	HBV	adw	"X"	1554	0
1.0803	TLQDPRVRGL	10	HBV	ayw	ENV	119	0
1.0804	NMENITSGFL	10	HBV	ayw	ENV	163	0
1.0891	DLVNLLPAIL	10	HCV		LORF	1878	0
1.0404	PLTEEKIKAL	10	HIV		POL	720	<0.0002
1.0409	QLGIPHPAGL	10	HIV		POL	786	<0.0002
1.0411	GLKKKKSVTV	10	HIV		POL	794	<0.0002
1.0450	PIWKGPAKLL	10	HIV		POL	1488	<0.0002
1.0476	DLAVAVEPVV	10	HCV		LORF	966	<0.0002
1.0478	SLTGRDKNQV	10	HCV		LORF	1046	<0.0002
1.0490	DLEVVTSTWV	10	HCV		LORF	1652	<0.0002
1.0494	GLGKVLIDIL	10	HCV		LORF	1843	<0.0002
1.0505	VLTTSCGNTL	10	HCV		LORF	2704	<0.0002
1.0506	ELITSCSSNV	10	HCV		LORF	2781	<0.0002

APPENDIX II: 10-MER PEPTIDES

Peptide	Sequence	AA	Virus	Strain	Molecule	Pos.	A2.1
1.0510	CLRKLGVPPL	10	HCV		LORF	2908	<0.0002
1.0511	PLGFFPDHQL	10	HBV	adr	ENV	10	<0.0002
1.0514	NMENTTSGFL	10	HBV	adr	ENV	163	<0.0002

Appendix III
PLP 8-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	Algorithm Score (EO2)
Hu PLP	10	8	C	L	V	G	A	P	F	A	
Hu PLP	13	8	G	A	P	F	A	S	L	V	
Hu PLP	23	8	C	L	C	F	F	G	V	A	
Hu PLP	39	8	A	L	T	G	T	E	K	L	
Hu PLP	40	8	L	T	G	T	E	K	L	I	
Hu PLP	60	8	Y	L	I	N	V	I	H	A	
Ms PLP	64	8	V	I	H	A	F	Q	C	V	
Hu PLP	64	8	V	I	H	A	F	Q	Y	V	
Hu PLP	74	8	G	T	A	S	F	F	F	L	
Hu PLP	80	8	F	L	Y	G	A	L	L	L	
Hu PLP	93	8	T	I	G	A	V	R	Q	I	
Hu PLP	106	8	T	T	I	C	G	K	G	L	
Hu PLP	131	8	Q	A	H	S	L	E	R	V	
Hu PLP	152	8	F	V	G	I	T	Y	A	L	
Hu PLP	154	8	G	I	T	Y	A	L	T	V	
Hu PLP	199	8	I	T	Y	A	L	T	V	V	
Hu PLP	157	8	Y	A	L	I	V	V	W	L	
Hu PLP	158	8	A	L	T	V	V	W	L	L	
Hu PLP	159	8	L	T	V	V	W	L	L	V	
Hu PLP	164	8	L	L	V	F	A	C	S	A	
Hu PLP	165	8	L	V	F	A	C	S	A	V	
Hu PLP	167	8	F	A	C	S	A	V	P	V	
Hu PLP	199	8	S	L	C	A	D	A	R	M	
Hu PLP	203	8	D	A	R	M	Y	G	V	L	
Hu PLP	212	8	W	I	A	F	F	G	R	V	
Hu PLP	218	8	K	V	C	G	S	N	L	L	
Hu PLP	224	8	L	L	S	I	C	K	T	A	
Hu PLP	234	8	Q	M	T	F	H	L	F	I	
Hu PLP	238	8	H	L	F	I	A	A	F	V	
Hu PLP	244	8	F	V	G	A	A	A	T	L	
Hu PLP	247	8	A	A	A	T	L	V	S	L	

Appendix III
PLP 8-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	Algorithm Score (EO2)
Hu PLP	248	8	A	A	T	L	V	S	L	L	
Hu PLP	253	8	S	L	L	T	F	M	I	A	
Hu PLP	254	8	L	L	T	F	M	I	A	A	
Hu PLP	260	8	A	A	T	Y	N	F	A	V	
Hu PLP	261	8	A	T	Y	N	F	A	V	L	

Appendix III
MBP 8-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	Algorithm Score (EO2)
Hu MBP	14	8	Y	L	A	T	A	S	T	M	
Hu MBP	34	8	D	T	G	I	L	D	S	I	
Hu MBP	65	8	R	T	A	H	Y	G	S	L	
Ms MBP	70	8	H	A	R	S	R	P	G	L	
Hu MBP	79	8	R	T	Q	D	E	N	P	V	
Hu MBP	86	8	V	V	H	F	F	K	N	I	
Ms MBP	87	8	R	T	T	H	Y	G	S	L	
Hu MBP	143	8	G	V	D	A	Q	G	T	L	
Hu MBP	149	8	T	L	S	K	I	F	K	L	

Appendix III
PLP 9-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	Algorithm Score (EO2)
Hu PLP	163	9	W	L	L	V	F	A	C	S	A	-18.67
Hu PLP	205	9	R	M	Y	G	V	L	P	W	I	-18.79
Hu PLP	145	9	W	L	G	H	P	D	K	F	V	-19.05
Hu PLP	253	9	S	L	L	T	F	M	I	A	A	-19.07
Hu PLP	251	9	L	V	S	L	L	T	F	M	I	-20.03
Hu PLP	258	9	M	I	A	A	T	Y	N	F	A	-20.32
Hu PLP	80	9	F	L	Y	G	A	L	L	L	A	-20.53
Ms PLP	205	9	R	M	Y	G	V	L	P	W	N	-20.69
Hu PLP	64	9	V	I	H	A	F	Q	Y	V	I	-20.71
Hu PLP	23	9	G	L	C	F	F	G	V	A	L	-21.23
Ms PLP	23	9	G	L	C	F	F	G	V	A	L	-21.23
Ms PLP	179	9	N	T	W	T	T	C	Q	S	I	-21.24
Hu PLP	233	9	F	Q	M	T	F	H	L	F	I	-21.25
Hu PLP	234	9	Q	M	T	F	H	L	F	I	A	-21.29
Hu PLP	259	9	I	A	A	T	Y	N	F	A	V	-21.32
Hu PLP	157	9	Y	A	L	T	V	V	W	L	L	-21.51
Hu PLP	76	9	A	S	F	F	F	L	Y	G	A	-21.52
Hu PLP	158	9	A	L	T	V	V	W	L	L	V	-21.56
Hu PLP	252	9	V	S	L	L	T	F	M	I	A	-21.58
Hu PLP	237	9	F	H	L	F	I	A	A	F	V	-21.61
Ms PLP	208	9	G	V	L	P	W	N	A	F	P	-21.61
Hu PLP	164	9	L	L	V	F	A	C	S	A	V	-21.81
Hu PLP	78	9	F	F	F	L	Y	G	A	L	L	-22.05
Hu PLP	250	9	T	L	V	S	L	L	T	F	M	-22.10
Hu PLP	208	9	G	V	L	P	W	I	A	F	P	-22.10
Hu PLP	39	9	A	L	T	G	T	E	K	L	I	-22.13
Hu PLP	240	9	F	I	A	A	F	V	G	A	A	-22.19
Hu PLP	235	9	M	T	F	H	L	F	I	A	A	-22.22
Hu PLP	244	9	F	V	G	A	A	A	T	L	V	-22.22
Ms PLP	64	9	V	I	H	A	F	Q	C	V	I	-22.33

Appendix III
PLP 9-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	Algorithm Score (EO2)
Hu PLP	12	9	V	C	A	P	F	A	S	L	V	-22.36
Hu PLP	45	9	K	L	I	E	T	Y	F	S	K	-22.42
Hu PLP	30	9	A	L	F	C	G	C	G	H	E	-22.46
Hu PLP	9	9	R	C	L	V	G	A	P	F	A	-22.52
Hu PLP	189	9	F	P	S	K	T	S	A	S	I	-22.54
Hu PLP	71	9	V	I	Y	G	T	A	S	F	F	-22.60
Hu PLP	73	9	Y	G	T	A	S	F	F	F	L	-22.63
Hu PLP	11	9	L	V	G	A	P	F	A	S	L	-22.64
Hu PLP	86	9	L	L	A	E	G	F	Y	T	T	-22.65
Ms PLP	63	9	N	V	I	H	A	F	Q	C	V	-22.65
Hu PLP	212	9	W	I	A	F	P	G	K	V	C	-22.67
Hu PLP	223	9	N	L	L	S	I	C	K	T	A	-22.68
Hu PLP	199	9	S	L	C	A	D	A	R	M	Y	-22.71
Hu PLP	179	9	N	T	W	T	T	C	D	S	I	-22.73
Hu PLP	201	9	G	A	D	A	R	M	Y	G	V	-22.74
Hu PLP	112	9	G	L	S	A	T	V	T	G	G	-22.78
Hu PLP	161	9	V	V	W	L	L	V	F	A	C	-22.78
Hu PLP	175	9	Y	I	Y	F	N	T	W	T	T	-22.81

Appendix III
PLP 9-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	Algorithm Score (EO2)
Hu PLP	56	9	Q	D	Y	E	Y	L	I	N	V	-22.84
Hu PLP	241	9	I	A	A	F	V	G	A	A	A	-22.87
Hu PLP	154	9	G	I	T	Y	A	L	T	V	V	-22.89
Hu PLP	257	9	F	M	I	A	A	T	Y	N	F	-22.89
Hu PLP	196	9	S	I	G	S	L	C	A	D	A	-22.90
Hu PLP	18	9	S	L	V	A	T	G	L	C	F	-22.91
Hu PLP	261	9	A	T	Y	N	F	A	V	L	K	-23.00
Hu PLP	171	9	A	V	P	V	Y	I	Y	F	N	-23.05
Hu PLP	70	9	Y	V	I	Y	G	T	A	S	F	-23.11
Hu PLP	22	9	T	G	L	C	F	F	G	V	A	-23.12
Hu PLP	134	9	S	L	F	R	V	C	H	C	L	-23.16
Hu PLP	16	9	F	A	S	L	V	A	T	G	L	-23.20
Hu PLP	74	9	G	T	A	S	F	F	F	L	Y	-23.20
Hu PLP	79	9	F	F	L	Y	G	A	L	L	L	-23.24
Hu PLP	246	9	G	A	A	A	T	L	V	S	L	-23.26
Hu PLP	181	9	W	T	T	C	D	S	I	A	F	-23.27
Hu PLP	28	9	G	V	A	L	F	C	G	C	G	-23.31
Hu PLP	247	9	A	A	A	T	L	V	S	L	L	-23.31
Hu PLP	219	9	V	C	G	S	N	L	L	S	I	-23.33
Hu PLP	160	9	T	V	V	W	L	L	V	F	A	-23.40
Hu PLP	54	9	N	Y	Q	D	Y	E	Y	L	I	-23.43
Hu PLP	107	9	T	I	C	G	K	G	L	S	A	-23.45
Hu PLP	166	9	V	F	A	C	S	A	V	P	V	-23.53
Hu PLP	2	9	G	L	L	E	C	C	A	R	C	-23.57
Hu PLP	167	9	F	A	C	S	A	V	P	V	Y	-23.60
Hu PLP	260	9	A	A	T	Y	N	F	A	V	L	-23.61
Hu PLP	152	9	F	V	G	I	T	Y	A	L	T	-23.63
Hu PLP	187	9	I	A	F	P	S	K	T	S	A	-23.63
Hu PLP	63	9	N	V	I	H	A	F	Q	Y	V	-23.65
Hu PLP	60	9	Y	L	I	N	V	I	H	A	F	-23.66
Hu PLP	85	9	L	L	L	A	E	G	F	Y	T	-23.66

Appendix III
PLP 9-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	Algorithm Score (EO2)
Ms PLP	210	9	L	P	W	N	A	F	P	G	K	-23.66
Hu PLP	198	9	G	S	L	C	A	D	A	R	M	-23.67
Hu PLP	20	9	V	A	T	G	L	C	F	F	G	-23.71
Hu PLP	263	9	Y	N	F	A	V	L	K	L	M	-23.71
Ms PLP	209	9	V	L	P	W	N	A	F	F	G	-23.71
Hu PLP	84	9	A	L	L	L	A	E	G	F	Y	-23.73
Hu PLP	206	9	M	Y	G	V	L	P	W	I	A	-23.77
Hu PLP	153	9	V	G	I	T	Y	A	L	T	V	-23.80
Hu PLP	269	9	K	L	M	G	R	G	T	K	F	-23.92
Hu PLP	138	9	V	C	H	C	L	G	K	W	L	-23.99
Hu PLP	3	9	L	L	E	C	C	A	R	C	L	-24.02
Hu PLP	92	9	Y	T	T	G	A	V	R	Q	I	-24.40
Hu PLP	21	9	A	T	G	L	C	F	F	G	V	-24.47
Hu PLP	192	9	K	T	S	A	S	I	G	S	L	-24.74
Hu PLP	38	9	E	A	L	T	G	T	E	K	L	-25.72
Hu PLP	105	9	K	T	T	I	C	G	K	G	L	-26.97

Appendix III
MBP 9-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	Algorithm Score (EO2)
Hu MBP	110	9	S	L	S	R	F	S	W	G	A	-21.42
Hu MBP	16	9	Y	L	A	T	A	S	T	M	D	-22.01
Ms MBP	59	9	W	L	K	Q	S	R	S	P	L	-22.60
Hu MBP	86	9	V	V	H	F	F	K	N	I	V	-22.80
Ms MBP	52	9	R	G	S	G	K	V	P	W	L	-22.87
Hu MBP	16	9	A	T	A	S	T	M	D	H	A	-23.11
Hu MBP	37	9	I	L	D	S	I	G	R	F	F	-23.11
Hu MBP	108	9	G	L	S	L	S	R	F	S	W	-23.34
Hu MBP	93	9	I	V	T	P	R	T	P	P	P	-23.11
Ms MBP	63	9	S	R	S	P	L	P	S	H	A	-23.47
Hu MBP	79	9	R	T	Q	D	E	N	P	V	V	-23.49
Hu MBP	129	9	G	R	A	S	D	Y	K	S	A	-23.53
Hu MBP	21	9	M	D	H	A	R	H	G	F	L	-23.60
Hu MBP	160	9	D	S	R	S	G	S	P	M	A	-23.63
Ms MBP	75	9	P	G	L	C	H	M	Y	K	D	-23.64
Hu MBP	112	9	S	R	F	S	W	G	A	E	G	-23.77
Hu MBP	160	9	R	S	G	S	P	M	A	R	R	-23.77
Hu MBP	159	9	R	D	S	R	S	G	S	P	M	-23.81
Hu MBP	85	9	P	V	V	H	P	F	K	N	I	-23.82
Hu MBP	136	9	S	A	H	K	G	F	K	G	V	-23.90
Hu MBP	149	9	T	L	S	K	T	F	K	L	G	-23.90
Ms MBP	162	9	K	G	F	K	G	A	Y	D	A	-23.92
Hu MBP	64	9	A	R	T	A	H	Y	G	S	L	-23.99
Ms MBP	166	9	G	A	Y	D	A	Q	G	T	L	-24.66
Hu MBP	148	9	G	T	L	S	K	I	F	K	L	-24.78
Hu MBP	145	9	D	A	Q	G	T	L	S	K	I	-25.25

Appendix III
PLP 10-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	Algorithm Score (EO2)
Ms PLP	178	10	F	N	T	W	T	T	C	Q	S	I	-24.68
Hu PLP	178	10	F	N	T	W	T	T	C	D	S	I	-25.14
Hu PLP	204	10	A	F	M	Y	G	V	L	P	W	I	-25.48
Hu PLP	163	10	W	L	L	V	F	A	C	S	A	V	-25.66
Hu PLP	218	10	K	V	C	G	S	N	L	L	S	I	-25.89
Hu PLP	250	10	T	L	V	S	L	L	T	F	M	I	-26.00
Hu PLP	19	10	L	V	A	T	G	L	C	F	F	G	-26.25
Hu PLP	78	10	F	F	F	L	Y	G	A	L	L	L	-26.68
Hu PLP	157	10	Y	A	L	T	V	V	W	L	L	V	-26.78
Hu PLP	84	10	A	L	L	L	A	E	G	F	Y	T	-26.77
Hu PLP	233	10	F	Q	M	T	F	H	L	F	I	A	-26.78
Hu PLP	80	10	F	L	Y	G	A	L	L	L	A	E	-26.79
Hu PLP	167	10	F	A	C	S	A	V	P	V	Y	I	-27.28
Hu PLP	165	10	L	V	F	A	C	S	A	V	P	V	-27.32
Hu PLP	4	10	L	E	C	C	A	R	C	L	V	G	-27.36
Hu PLP	253	10	S	L	L	T	F	M	I	A	A	T	-27.42
Hu PLP	135	10	L	E	R	V	C	H	C	L	G	K	-27.48
Hu PLP	176	10	I	Y	F	N	T	W	T	T	C	D	-27.62
Hu PLP	24	10	L	C	F	F	G	V	A	L	F	C	-27.74
Hu PLP	146	10	L	G	H	P	D	K	F	V	G	I	-27.88
Hu PLP	237	10	F	H	L	F	I	A	A	F	V	G	-27.95
Hu PLP	56	10	Q	D	Y	E	Y	L	I	N	V	I	-27.99
Ms PLP	204	10	A	R	M	Y	G	V	L	P	W	N	-28.01
Hu PLP	158	10	A	L	T	V	V	W	L	L	V	F	-28.04
Hu PLP	137	10	R	V	C	H	C	L	G	K	W	L	-28.15
Hu PLP	72	10	I	Y	G	T	A	S	F	F	F	L	-28.16
Hu PLP	63	10	N	V	I	H	A	F	Q	Y	V	I	-28.17
Hu PLP	208	10	G	V	L	P	W	I	A	F	P	G	-28.17
Hu PLP	27	10	F	G	V	A	L	F	C	G	C	G	-28.29
Hu PLP	85	10	L	L	L	A	E	G	F	Y	T	T	-28.32
Ms PLP	62	10	I	N	V	I	H	A	F	Q	C	V	-28.33

Appendix III
PLP 10-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	Algorithm Score (EO2)
Hu PLP	222	10	S	N	L	L	S	I	C	K	T	A	-28.50
Hu PLP	76	10	A	S	F	F	F	L	Y	G	A	L	-28.43
Ms PLP	208	10	G	V	L	P	W	N	A	F	P	G	-28.45
Hu PLP	207	10	Y	G	V	L	F	W	I	A	F	P	-28.46
Hu PLP	79	10	F	F	L	Y	G	A	L	L	L	A	-28.49
Hu PLP	236	10	T	F	H	L	F	I	A	A	F	V	-28.50
Hu PLP	240	10	F	I	A	A	F	V	G	A	A	A	-28.51
Hu PLP	181	10	W	T	T	C	D	S	I	A	F	P	-28.56
Hu PLP	224	10	L	L	S	I	C	K	T	A	E	F	-28.56
Hu PLP	10	10	C	L	V	G	A	P	F	A	S	L	-28.62
Hu PLP	152	10	F	V	G	I	T	Y	A	L	T	V	-28.64
Hu PLP	62	10	I	N	V	I	H	A	F	Q	Y	V	-28.64
Hu PLP	214	10	A	F	F	G	K	V	C	G	S	N	-28.65
Hu PLP	188	10	A	F	P	S	K	T	S	A	S	I	-28.65
Hu PLP	99	10	Q	I	F	G	D	Y	K	T	T	I	-28.69
Hu PLP	18	10	S	L	V	A	T	G	L	C	F	F	-28.73
Hu PLP	3	10	L	L	E	C	C	A	R	C	L	V	-28.75

Appendix III
PLP 10-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	Algorithm Score (EO2)
Hu PLP	17	10	A	S	L	V	A	T	G	L	C	F	-28.76
Hu PLP	144	10	K	W	L	G	H	P	D	K	F	V	-28.78
Ms PLP	181	10	W	T	T	C	Q	S	I	A	F	P	-28.78
Hu PLP	159	10	L	T	V	V	W	L	L	V	F	A	-28.79
Hu PLP	174	10	V	Y	I	Y	F	N	T	W	T	T	-28.80
Hu PLP	248	10	A	A	T	L	V	S	L	L	T	F	-28.84
Hu PLP	23	10	G	L	C	F	F	G	V	A	L	F	-28.87
Hu PLP	209	10	V	L	P	W	I	A	F	P	G	K	-28.87
Hu PLP	29	10	V	A	L	F	C	G	C	G	H	E	-28.90
Hu PLP	261	10	A	T	Y	N	F	A	V	L	K	L	-28.92
Ms PLP	63	10	N	V	I	N	A	F	Q	C	V	I	-28.93
Hu PLP	74	10	G	T	A	S	F	F	F	L	Y	G	-28.93
Hu PLP	259	10	I	A	A	T	Y	N	F	A	V	L	-29.06
Hu PLP	242	10	A	A	F	V	G	A	A	A	T	L	-29.06
Hu PLP	2	10	G	L	L	E	C	C	A	R	C	L	-29.30
Hu PLP	257	10	F	M	I	A	A	T	Y	N	F	A	-29.37
Hu PLP	20	10	V	A	T	G	L	C	F	F	G	V	-29.41
Ms PLP	205	10	R	M	Y	G	V	L	P	W	N	A	-29.43
Hu PLP	155	10	I	T	Y	A	L	T	V	V	W	L	-29.60
Hu PLP	30	10	A	L	F	C	G	C	G	H	E	A	-29.70
Hu PLP	205	10	R	M	Y	G	V	L	F	W	I	A	-29.74
Hu PLP	258	10	M	I	A	A	T	Y	N	F	A	V	-30.06
Hu PLP	238	10	Q	M	T	F	H	L	F	I	A	A	-30.29
Hu PLP	238	10	H	L	F	I	A	A	F	V	G	A	-30.64
Hu PLP	246	10	G	A	A	A	T	L	V	S	L	L	-30.64
Hu PLP	38	10	E	A	L	T	G	T	E	K	L	I	-30.92
Hu PLP	230	10	T	A	E	F	Q	M	T	F	H	L	-31.03
Hu PLP	11	10	L	V	G	A	P	F	A	S	L	V	-31.25
Hu PLP	201	10	C	A	D	A	R	M	Y	G	V	L	-31.73

Appendix III
PLP 11-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	11	Algorithm Score (EO2)
Hu PLP	2	11	G	L	L	E	C	C	A	R	C	L	V	
Hu PLP	10	11	C	L	V	G	A	P	F	A	S	L	V	
Hu PLP	19	11	L	V	A	T	G	L	C	F	F	G	V	
Hu PLP	21	11	A	T	G	L	C	P	P	G	V	A	L	
Hu PLP	30	11	A	L	F	C	G	C	G	H	E	A	L	
Hu PLP	61	11	L	I	N	V	I	H	A	F	Q	Y	V	
Ms PLP	61	11	L	I	N	V	I	H	A	F	Q	C	V	
Hu PLP	71	11	V	I	Y	G	T	A	S	P	F	F	L	
Hu PLP	75	11	T	A	S	F	F	F	L	Y	G	A	L	
Hu PLP	86	11	L	L	A	E	G	P	Y	T	T	G	A	
Hu PLP	87	11	L	A	E	G	P	Y	T	T	G	A	V	
Hu PLP	107	11	T	I	C	G	K	G	L	S	A	T	V	
Hu PLP	145	11	W	L	G	H	P	D	K	P	V	G	I	
Hu PLP	152	11	F	V	G	I	T	Y	A	L	T	V	V	
Hu PLP	154	11	G	I	T	Y	A	L	T	V	V	W	L	
Hu PLP	155	11	I	T	Y	A	L	T	V	V	W	L	L	
Hu PLP	158	11	A	L	T	V	V	W	L	L	V	F	A	
Hu PLP	164	11	L	L	V	F	A	C	S	A	V	P	V	
Hu PLP	187	11	I	A	F	P	S	K	T	S	A	S	I	
Hu PLP	199	11	S	L	C	A	D	A	R	M	Y	G	V	
Hu PLP	203	11	D	A	R	M	Y	G	V	L	P	W	I	
Hu PLP	209	11	V	L	F	W	I	A	P	P	G	K	V	
Ms PLP	209	11	V	L	F	W	N	A	F	P	G	K	V	
Hu PLP	229	11	K	T	A	E	P	Q	M	T	F	H	L	
Hu PLP	235	11	M	T	F	H	L	F	I	A	A	F	V	
Hu PLP	238	11	H	L	F	I	A	A	F	V	G	A	A	
Hu PLP	241	11	I	A	A	F	V	G	A	A	A	T	L	
Hu PLP	242	11	A	A	F	V	G	A	A	A	T	L	V	
Hu PLP	244	11	F	V	G	A	A	A	T	L	V	S	L	
Hu PLP	249	11	A	T	L	V	S	L	L	T	F	M	I	

Appendix III
PLP 11-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	11	Algorithm Score (EO2)
Hu PLP	250	11	T	L	V	S	L	L	T	F	M	I	A	
Hu PLP	257	11	F	M	I	A	A	T	Y	N	F	A	V	
Hu PLP	258	11	M	I	A	A	T	Y	N	F	A	V	L	
Hu PLP	260	11	A	A	T	Y	N	F	A	V	L	K	L	

Appendix III
MBP 10-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	Algorithm Score (EO2)
Hu MBP	37	10	I	L	D	S	I	G	R	F	F	G	-27.66
Hu MBP	28	10	F	L	P	R	H	R	D	T	G	I	-27.85
Ms MBP	167	10	A	Y	D	A	Q	G	T	L	S	K	-28.54
Hu MBP	89	10	F	F	K	N	I	V	T	P	R	T	-28.98
Hu MBP	14	10	Y	L	A	T	A	S	T	M	D	H	-28.75
Hu MBP	84	10	N	P	V	V	H	F	F	K	N	I	-28.80
Hu MBP	32	10	H	R	D	T	G	I	L	D	S	I	-28.83
Hu MBP	110	10	S	L	S	R	F	S	W	G	A	E	-28.98
Hu MBP	85	10	P	V	V	H	F	F	K	N	I	V	-30.82
Ms MBP	85	10	H	T	R	T	T	H	Y	G	S	L	-31.29
Hu MBP	20	10	T	M	D	H	A	R	H	G	F	L	-31.40
Hu MBP	63	10	P	A	R	T	A	H	Y	G	S	L	-31.76
Ms MBP	48	10	G	A	P	K	R	G	S	G	K	V	-32.21

Appendix III
MBP 11-mers

Source	Peptide	AA	1	2	3	4	5	6	7	8	9	10	11	Algorithm Score (EO2)
Hu MBP	14	11	Y	L	A	T	A	S	T	M	D	H	A	
Hu MBP	19	11	S	T	M	D	H	A	R	H	G	F	L	
Hu MBP	28	11	F	L	P	R	H	R	D	T	G	I	L	
Hu MBP	108	11	G	L	S	L	S	R	F	S	W	G	A	
Hu MBP	143	11	G	V	D	A	Q	G	T	L	S	K	I	

WHAT IS CLAIMED IS:

1. A composition comprising an immunogenic peptide having an HLA-A2.1 binding motif, which immunogenic peptide has 9 residues and the following residues:

a first conserved residue at the second position from the N-terminus selected from the group consisting of I, V, A and T;

a second conserved residue at the C-terminal position selected from the group consisting of V, L, I, A and M.

2. A composition comprising an immunogenic peptide having an HLA-A2.1 binding motif, which immunogenic peptide has 9 residues:

a first conserved residue at the second position from the N-terminus selected from the group consisting of L, M, I, V, A and T;

a second conserved residue at the C-terminal position selected from the group consisting of A and M;

3. The composition of claim 1, wherein the amino acid at position 1 is not an amino acid selected from the group consisting of D, and P.

4. The composition of claim 2, wherein the amino acid at position 1 is not an amino acid selected from the group consisting of D, and P.

5. The composition of claim 1, wherein the amino acid at position 3 from the N-terminus is not an amino acid selected from the group consisting of D, E, R, K and H.

6. The composition of claim 2, wherein the amino acid at position 3 from the N-terminus is not an amino acid selected from the group consisting of D, E, R, K and H

7. The composition of claim 1, wherein the amino acid at position 6 from the N-terminus is not an amino acid selected from the group consisting of R, K and H.

8. The composition of claim 2, wherein the amino acid at position 6 from the N-terminus is not an amino acid selected from the group consisting of R, K and H.

9. The composition of claim 1, wherein the amino acid at position 7 from the N-terminus is not an amino acid selected from the group consisting of R, K, H, D and E.

10. The composition of claim 2, wherein the amino acid at position 7 from the N-terminus is not an amino acid selected from the group consisting of R, K, H, D and E.

11. A composition comprising an immunogenic peptide having an HLA-A2.1 binding motif, which immunogenic peptide has about 10 residues:

a first conserved residue at the second position from the N-terminus selected from the group consisting of L, M, I, V, A, and T; and

a second conserved residue at the C-terminal position selected from the group consisting of V, I, L, A and M;

wherein the first and second conserved residues are separated by 7 residues.

12. The composition of claim 11, wherein the amino acid at position 1 is not an amino acid selected from the group consisting of D, E and P.

13. The composition of claim 11, wherein the amino acid at position 3 from the N-terminus is not an amino acid selected from the group consisting of D and E.

14. The composition of claim 11, wherein the amino acid at position 4 from the N-terminus is not an amino acid selected from the group consisting of A, K, R and H.

15. The composition of claim 11, wherein the amino acid at positon 5 from the N-terminus is not P.

16. The composition of claim 11, wherein the amino acid at position 7 from the N-terminus is not an amino acid selected from the group consisting of R, K and H.

17. The composition of claim 11, wherein the amino acid at position 8 from the N-terminus is not an amino acid selected from the group consisting of D, E, R, K and H.

18. The composition of claim 11, wherein the amino acid at position 9 from the N-terminus is not an amino acid selected from the group consisting of R, K and H.

19. A pharmaceutical composition comprising a pahramceutically acceptable carrier and a therapeutically effective amount of a peptide capable of binding an HLA-A2.1 moelcule and inducing an immune response in a mammal.

20. The pharmaceutical composition of claim 19, wherein the peptide has a formula as follows: TLGIVCPI.

21. The pharamceutical composition of claim 19, further comprising a peptide having a formula as follows: YMLDLQPETT.

22. The pharmaceutical composition of claim 19, further comprising a T helper peptide.

23. The pharmaceutical composition of claim 22, wherein the T helper peptide has a formula as follows: aKXVAAWTLKAAa, wherein a is D-alanine and X is cyclohexylalanine.

HLA-A PURIFICATION AND
PEPTIDE ELUTION

CELLULAR SOURCE: HLA-ANTIGENS ($5-10 \times 10^9$ CELL EQUIVALENTS)	A) EBV TRANSFORMED B CELL LINES - HOMOZYGOUS B) HLA-A TRANSFECTANTS - e.g. .221-HLA-A1 C) P815 TRANSFECTANTS (MOUSE MASTOCYTOMA)
↓	
DETERGENT LYSIS (10^8 CELLS/ml)	1% NP-40 OR 1% RENEX 30 PLUS PROTEASE INHIBITORS - 1 HR, 4°C
↓	
DETERGENT LYSATE	CENTRIFUGATION AT 15,000 x g, 30 MIN.
↓	
AFFINITY CHROMATOGRAPHY	mAb-SEPHAROSE 5 mg/ml 5-10 ml COLUMN
↓	
PURIFIED HLA-A ANTIGEN	ANTICIPATED YIELDS 450-900 µg
↓	
ACID TREATMENT	10% ACETIC ACID, 5 MIN, 100°C
↓	
PEPTIDES	YM3 FILTRATION, 3kD CUT-OFF
↓	
SEQUENCE/MOTIF	D. HUNT - HPLC/EI-TMS CYTEL - HPLC/ABI 477A

FIG. 1.

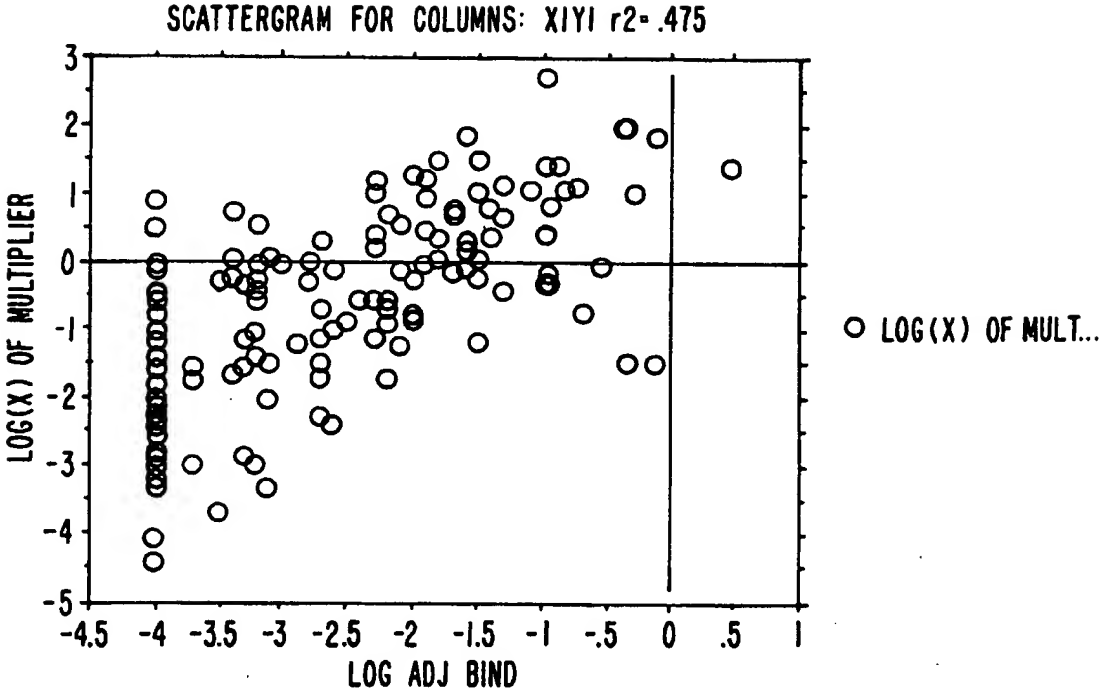

SCATTERGRAM FOR COLUMNS: XIYI r2= .475
LOG(X) OF MULTIPLIER
LOG ADJ BIND
O LOG(X) OF MULT...

FIG. 2.

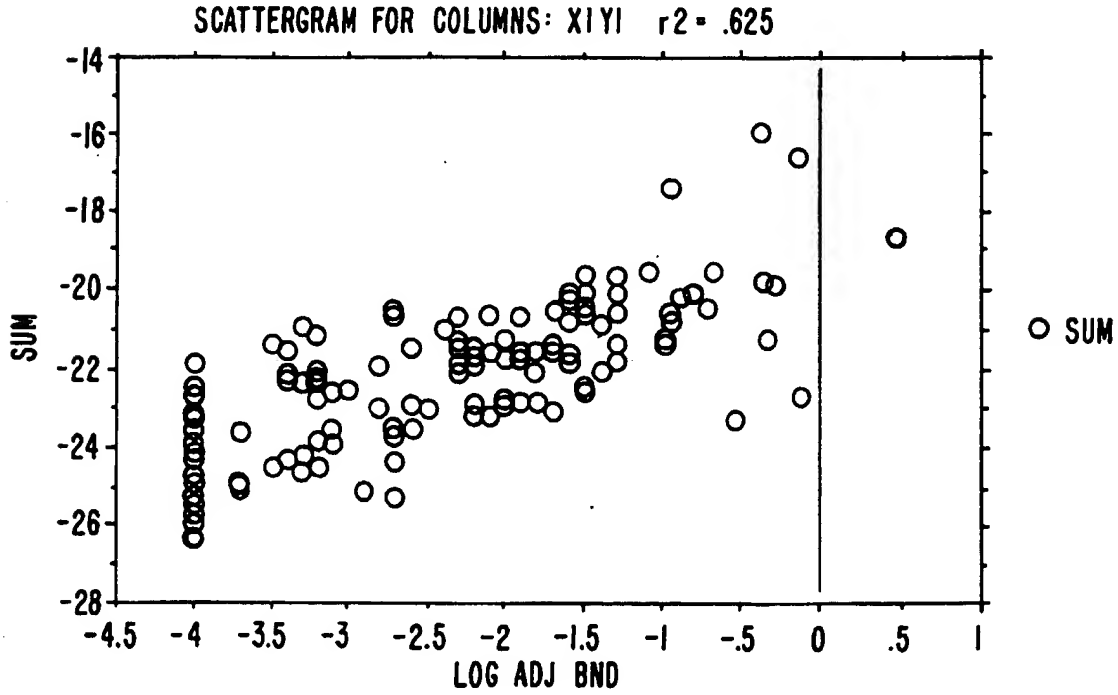

SCATTERGRAM FOR COLUMNS: XIYI r2 = .625
SUM
LOG ADJ BND
O SUM

FIG. 3.

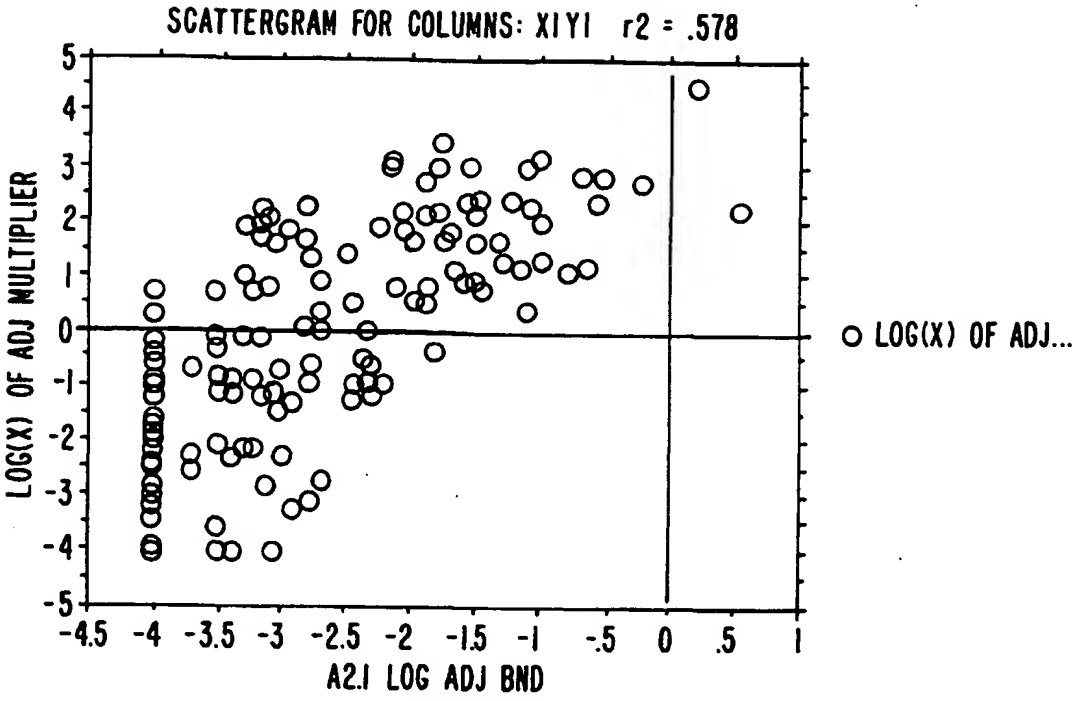

SCATTERGRAM FOR COLUMNS: X1Y1 r2 = .578
LOG(X) OF ADJ MULTIPLIER
A2.1 LOG ADJ BND
O LOG(X) OF ADJ...

FIG. 4.

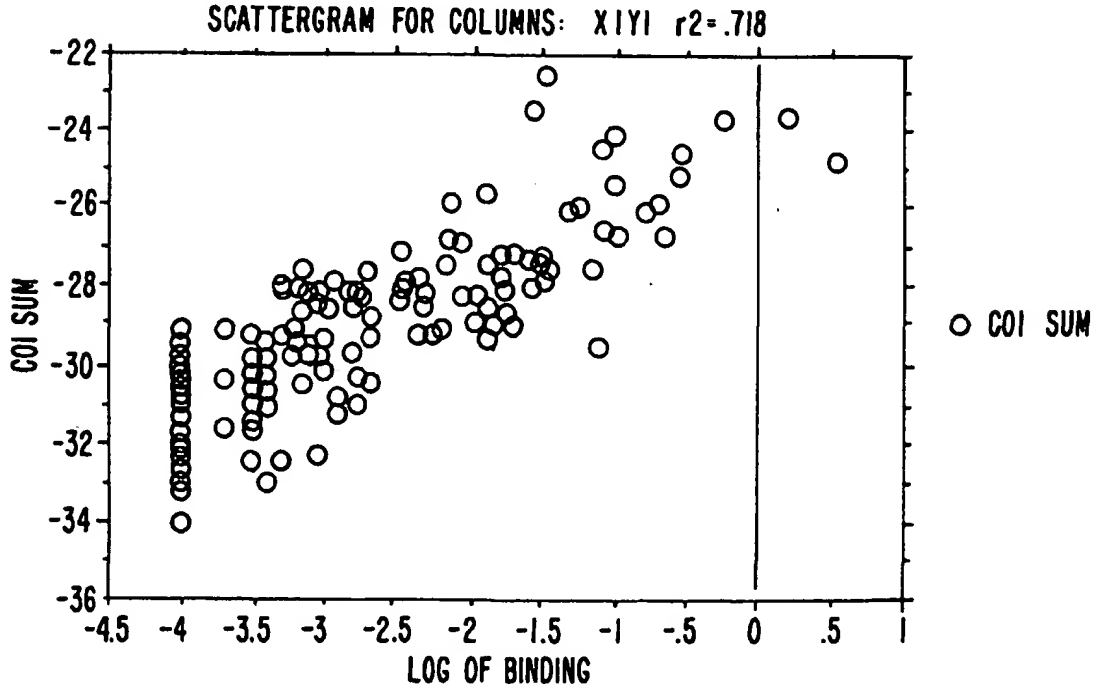

SCATTERGRAM FOR COLUMNS: X1Y1 r2=.718
COI SUM
LOG OF BINDING
O COI SUM

FIG. 5.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/02353

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :A61K 37/02; C07K 7/06
US CL :424/88; 530/328
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/88; 530/328, 868; 514/885

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Medline, CAS Registry

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	Journal of Immunolgy, Volume 147, No. 11, issued 01 December 1991, Sette et al, "Random Association Between the Peptide Repertiore of A2.1 Class I and Several Different DR Class II Molecules", pages 3893, see page 3897-3900, Table III.	2, 4, 6, 10, 19 ---------- 22
X	Science, Volume 255, issued 06 March 1992, Henderson et al, "HLA- A2.1 Associated Peptides from a Mutant Cell Line: A Second Pathway of Antigen Presentation", pages 1264-1266, see page 1265.	11-18
Y	Nature, Volume 351, issued 23 May 1991, Falk et al, "Allele-specific Motifs Revealed by Sequencing of Self-Peptides Eluted from MHC Molecules", pages 290-296, see page 293.	1, 3, 5, 7, and 9

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 29 APRIL 1994	Date of mailing of the international search report 24 MAY 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JULIE KRSEK-STAPLES Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

Ir tional application No.
PCT/US94/02353

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	A. Lehninger, "Principles of Biochemistry" published 1982 by Worth Publishers, Inc. (N.Y.), pages 100-101, see pages 100-101.	1, 3, 5, 7, and 9
Y	European Journal of Immunology, Volume 21, issued June 1991, Sarobe et al, "Induction of Antibodies Against a Peptide Hapten Does Not Require Covalent Linkage Between the Hapten and a Class II Presentable T Helper Peptide", pages 1555-1558, see page 1556.	22

PAGE4 Is a Cytoplasmic Protein That Is Expressed in Normal Prostate and in Prostate Cancers

Carlo Iavarone, Curt Wolfgang,[1] Vasantha Kumar, Paul Duray, Mark Willingham, Ira Pastan, and Tapan K. Bera[2]

Laboratory of Molecular Biology [C. I., C. W., V. K., I. P., T. K. B.], and of Pathology [P. D.], Clinical Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, and Department of Pathology, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157 [M. W.]

Abstract

***PAGE4* is an X chromosome-linked cancer-testis antigen that was identified by expressed sequence tags database mining and a functional genomic approach. *PAGE4* is preferentially expressed in normal male and female reproductive tissues and also in a variety of cancers including prostate. In the present study, we have used *in situ* hybridization to show that *PAGE4* mRNA is expressed only in the epithelial cells of normal and prostate-cancer specimens. Analysis of the protein product encoded by the *PAGE4* mRNA reveals that it encodes a M_r 16,000 protein and is detected in tissue extracts from both normal prostate and prostate cancer. Cell fractionation analysis of PAGE4 protein indicates that PAGE4 is localized in the cytoplasm of the cell. Furthermore, cDNA microarray analysis indicates that the expression of *lipoprotein lipase*, a gene frequently deleted in prostate cancer, is down-regulated in a cell line that expresses *PAGE4*.**

Introduction

Prostate cancer is a major public health problem and the second leading cause of death for men in the United States. About one in five men in the United States will develop prostate cancer during their lifetime. Despite its distinction as the most frequently diagnosed noncutaneous cancer, little is known about the causes of this disease largely because of the cellular heterogeneity of the prostate and the lack of systematic analysis of the genes expressed in this tissue. Completion of the human genome project and the technological advancement in biomedical research has enabled researchers to identify and to systematically analyze new genes that could be used as targets for cancer therapy or that could be involved in the multistep process of cancer. Many different methods are currently being used to identify tissue- or cancer-specific genes. Our laboratory is interested in identifying genes by using the EST[3] database, and developed a computer-based screening strategy to generate clusters of ESTs that are specifically expressed in normal prostate and/or prostate cancer but not in essential normal tissues (1). We then used experimental approaches to verify these predictions. Using this approach, we identified *PAGE4*, a new gene expressed in normal prostate and testis and highly expressed in prostate and uterine cancers (2).

In this report, we analyzed *PAGE4* mRNA expression by *in situ* hybridization using several prostate cancer samples. We also report here the characterization and subcellular localization of the protein encoded by *PAGE4* mRNA. Using cell fractionation and immunofluorescence analysis we demonstrate that the M_r 16,000 protein product of *PAGE4* is localized in the cytoplasm of the cell. In addition, using cDNA microarray analysis, we report that the expression of *LPL*, a gene involved in lipid metabolism, is down-regulated in cells expressing *PAGE4*.

Materials and Methods

Primers. Nucleotide sequences of the primers used in this study were: forward CP1 (5′-AAGAGGAATTCGCAGCGATGAGTGCACGA-3′), and reverse CP2 (5′-GCACTGAATTCTCAGCCATGTGTGTAGCT-′); forward CR1 (5′-AAGAGCATATGAGTGCACGAAGTGAGTCA-′), and reverse CR2 (5′-CACTCTCGAGTGGCTGCCCATCTCCTGCTTC-3prime]); and forward LU1 (5′-TGCATGTGTGTTGTCCTTCAG-3′), and reverse LU2 (5′-GAGTTCCCAGTGCTATGTGCTGCT-3′). All of the primers were synthesized by Genosys (The Woodlands, TX).

Constructs. The pCI-*PAGE4* plasmid was constructed from the pLZR-SpBMN-Z vector (3) replacing the *lacZ* gene with the *PAGE4* cDNA. The DNA fragment encoding the open reading frame of *PAGE4* was amplified from the *PAGE4* cDNA using primer pairs CP1 and CP2. The PCR product was gel purified, digested with *Eco*RI, and ligated into an *Eco*RI-digested pLZR-SpBMN-Z vector. The pET-*PAGE4* plasmid was constructed using the pET23a plasmid (Novagen, Madison, WI). The DNA fragment encoding the open reading frame of *PAGE4* was amplified from the *PAGE4* cDNA using primer pairs CR1 and CR2. The PCR product was gel purified, digested with *Nde*I and *Xho*I, and ligated into a *Nde*I-*Xho*I-digested pET23a vector.

The pBS-*PAGE4* plasmid was constructed using the pBluescript II SK (+) plasmid (Stratagene, La Jolla, CA). The pBS-*PAGE4* plasmid contains the nucleotides 1–442 of the *PAGE4* transcript.

Received 11/23/01; revised 1/28/02; accepted 1/28/02.

[1] Present address: Novartis Pharmaceuticals Corporation, 9 West Watkins Mill Road, Gaithersburg, Maryland 20878.

[2] To whom reprint requests should be addressed, at Laboratory of Molecular Biology, Clinical Cancer Research, National Cancer Institute, NIH, 37 Convent Drive MSC 4264, Building 37, Room 5106, Bethesda, MD 20892-4264. Phone: (301) 496-0976; Fax: (301) 402-1344; E-mail: tkbera@helix.nih.gov.

[3] The abbreviations used are: EST, expressed sequence tag; CT, cancer-testis; Dlk1, δ-like homologue *Drosophila*; IGFbp-2, insulin-like growth factor binding protein 2; LPL, lipoprotein lipase; RT-PCR, reverse transcription-PCR; NCI, National Cancer Institute; CCR, Clinical Cancer Research.

EXHIBIT B

Cell Lines and Culture. NIH3T3 and 293T cells were maintained at 37°C in 5% CO_2 in DMEM supplemented with 10% fetal bovine serum (Quality Biological, Inc., Gaithersburg, MD), 2 mM L-glutamine, and 1 mM penicillin-streptomycin. PC3 cells were maintained at 37°C in 5% CO_2 in RPMI medium supplemented with 10% fetal bovine serum (Quality Biological), 2 mM L-glutamine, 1 mM sodium pyruvate and penicillin-streptomycin.

Packaging Cells Transfection and Target Cells Infection. The amphotropic 293T packaging line was transfected by a calcium-phosphate/chloroquine method described previously (3). Culture supernatants containing viral particles were collected 48 h after transfection, filtered with a 0.22 μm filter unit (Millipore, Bedford, MA), and used to infect the target cells. For stable infections, cells were plated in dishes of 100 mm at low density (300 cells/dish) and medium density (1000 cells/dish) 48 h after the infection. Three days after the infection, NIH3T3 cells were selected with 2.5 μg/ml puromycin, and PC3 cells were selected with 0.5 μg/ml puromycin. After 15 days, individual clones were picked and grown in the absence of puromycin.

Comparative cDNA Microarray Analysis. The protocol used for microarray analysis was designed by the Microarray Core Facility at NCI and can be found on the NCI/CCR microarray homepage.[4] Mm-OC-6.1p mouse array chips were obtained from the NCI microarray core facility. After hybridization, arrays were scanned using an Axon GenePix 4000 scanner and processed using the GenePix software. The results were analyzed using tools found on the NCI/CCR microarray homepage.

RNA Isolation for Microarray Analysis. Total RNA was isolated from the clones infected with pCI-*PAGE4*, pLZR-SpBMN-Z, and the empty vector. The RNA was prepared by using TRIzol Reagent according to the manufacturer's instructions (Life Technologies, Inc., Rockville, MD) with the following modifications: after the chloroform addition and the phase separation, the aqueous phase was used in a second round of purification using the Rneasy Maxi Kit (Qiagen, Chatsworth, CA) as recommended by the manufacturer.

mRNA Extraction and Northern Analysis. mRNA from each cell line was isolated using a FastTrack kit (Invitrogen, Carlsbad, CA). Two μg mRNA per lane were electrophoresed under denaturing conditions and subsequently transferred to a nylon membrane according to established procedure. The *PAGE4* cDNA fragment was labeled with ^{32}P by random primer extension (Lofstrand Labs Ltd., Gaithersburg, MD). Hybridization was performed as described previously (4).

Dot Blot Analysis of Matched Tumor/Normal Expression Array. A membrane with 68 separate samples of cDNA synthesized from human tumors and corresponding normal tissue from the same individual (Clontech, Palo Alto, CA) was hybridized with a ^{32}P-labeled *PAGE4* cDNA fragment (Lofstrand Labs Ltd.). Hybridization conditions were described previously (4).

RT-PCR. For the analysis of LPL expression on PC3 and prostate mRNA, RT-PCR was performed as described previously (4) using primer pair LU1 and LU2.

***In Situ* Hybridization.** The paraffin-embedded prostate tissue sections were deparaffinized by placing the slides over a slide warmer at 65°C for 1 h. The slides were then rinsed in two changes of xylene for 5 min each and air-dried. They were then rinsed in two changes of absolute alcohol for 5 min and air-dried. Biotinylated probes were prepared using *PAGE4* (442 bp) and U6 (250 bp) cDNA cloned in pBluescript II (+) plasmid. Biotinylated pBluescript II (+) without any insert was used as negative control. Probes were labeled using the BioNick Labeling System kit (Life Technologies, Inc., Rockville, MD) according to the manufacturer' s instructions. The probes were incubated at 16°C for 3 h. The unincorporated nucleotides from the labeled DNA probe were removed by three ethanol precipitations. Slides were hybridized using the *In Situ* Hybridization and Detection system (Life Technologies, Inc.) according to the manufacturer' s instructions. The slides were counterstained using 0.2% Light Green stain, rinsed through a series of alcohol grades, and mounted in Cytoseal. Microscopic evaluation (brightfield) was performed using a Nikon Eclipse 800 microscope (5).

Preparation of Cell Extracts. Protein extracts from different cell lines were prepared as described previously (6). Briefly, about 5 × 10^6 growing cells (80% confluent) from each respective cell line were harvested and resuspended in 1× RIPA buffer containing proteinase inhibitors [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 1 mM phenylmethylsulfonyl fluoride, 10 μg/ml aprotinin, and 10 μg/ml leupeptin]. The extracts were rotated for 30 min at 4°C and clarified by centrifugation. Protein concentrations were determined by using the Coomassie Plus Protein Assay reagent according to the manufacturer's instructions (Pierce, Rockport, IL).

Protein extracts from normal prostate and prostate cancer tissue were prepared by grinding 0.5 g of tissue frozen at −80°C into a fine powder using a cold mortar and pestle. The powdered tissue was collected, resuspended in 1× RIPA buffer, and sonicated briefly and clarified by centrifugation.

Nuclear, membrane, and cytoplasmic extracts from NIH3T3 and PC3 cells, infected with *PAGE4*, were prepared based on published protocols (7, 8).

Preparation of Recombinant PAGE4 Protein. The plasmid pET-*PAGE4* encodes amino acids 1 to 102 of *PAGE4* with six histidines at the COOH terminus encoded by the vector to facilitate purification of the protein. The recombinant PAGE4 protein was then expressed in *Escherichia coli* and purified using Ni^{2+}NTA agarose matrix following the supplier's instructions (Qiagen Inc.).

Production of Polyclonal Anti-PAGE4 Antibodies in Rabbits and Purification of IgG from Antisera. Purified PAGE4 protein was diluted to 0.5 mg/ml and injected into rabbits with complete Freund's adjuvant for the first immunization, and with incomplete Freund's adjuvant for subsequent immunizations. For PAGE4 antipeptide antibody, a peptide of 15 amino acids (amino acids 46–63) was synthesized and then injected into rabbits as described above. Sera

[4] Internet address: http://nciarray.nci.nih.gov/.

were collected after the fourth, fifth, and sixth immunizations and titrated by ELISA against the purified recombinant PAGE4 protein. Total IgG was then purified with immobilized protein A (Pierce) matrix following the supplier's instructions.

For tissue analysis, the *PAGE4* antisera were further purified by using immobilized *E. coli* lysate kit (Pierce) according to the manufacturer's instructions. The *PAGE4*-peptide antisera were further affinity-purified by using a HiTrap *N*-hydroxy-succinimide-activated column coupled with the recombinant *PAGE4*, according to the manufacturer's instructions.

Western Blot Analysis. Ten μg of protein extract from cell lines and 80 μg from tissues were run on a 16.5% Tris-Tricine gel (BIO-RAD, Hercules, CA) and transferred to a 0.2 μm Immun-Blot polyvinylidine difluoride membrane (BIO-RAD) in transfer buffer [25 mM Tris, 192 mM glycine, 20% (v/v) methanol (pH 8.3)] at 4°C for 2 h at 50 V. Filters were probed with 10 μg/ml protein A-purified anti-*PAGE4* antiserum or 1 μg/ml affinity column-purified anti-*PAGE4*-peptide antiserum, and their respective signals were detected using a chemiluminescence Western blotting kit according to the manufacturer's instructions (Roche Molecular Biochemicals, Indianapolis, IN).

Immunofluorescence Analysis. Immunofluorescence analysis was performed as previously described (9). Briefly, 293T and NIH3T3 cells, transfected with pCI-*PAGE4* plasmid, were grown in 35-mm dishes. Cells were then fixed in 3.7% formaldehyde in PBS for 15 min at 23°C and then washed in PBS. All of the subsequent incubations (at 23°C) included 0.1% saponin (Sigma Chemical Co., St. Louis, MO) and 1% BSA in PBS (BSA-sap-PBS). Fixed cells were first incubated in BSA-sap-PBS for 30 min and then incubated with affinity-purified anti-PAGE4 antibodies (10 μg/ml) for 30 min. After washing in PBS, the cells were incubated with affinity-purified goat antirabbit IgG (H+L) conjugated to rhodamine (25 μg/ml) in BSA-PBS for 30 min. After washing, cells were fixed in 3.7% formaldehyde and mounted under coverslips in buffered glycerol. Microscopic evaluation of cells expressing PAGE4 protein was performed by direct observation of cells using a Zeiss Axiplan2 fluorescent microscope equipped with rhodamine filters, and images were captured using a Dage 300 cooled-chip charge coupled device camera.

Results

Expression of *PAGE4* in Various Cancers. By using Northern blot and RT-PCR analysis, we previously reported that *PAGE4* is expressed in normal prostate, prostate cancer, and female reproductive tissues (2). To determine whether *PAGE4* is expressed in other types of cancer, we conducted a cDNA dot blot analysis using a cancer-profiling array (Clontech). This array contains samples from 68 different cancers and their corresponding normal tissues. As shown in Fig. 1, among the 68 different samples from human normal and cancer tissues, *PAGE4* was detected in all three of the normal prostate samples (D11, 12, and 13) and prostate cancer samples (E11, 12, and 13). Moreover, *PAGE4* was found in one normal cervix sample and cervical cancer sample (G14 and H14, respectively). *PAGE4* was also detected in four uterine cancer samples (H1, 3, 4, and 5), three normal uterus samples (G3, 4, and 5), and one ovary sample (G10). Interestingly, the expression of *PAGE4* was also found in two kidney cancer samples (B2 and 5) but not in normal kidney (A1–A11).

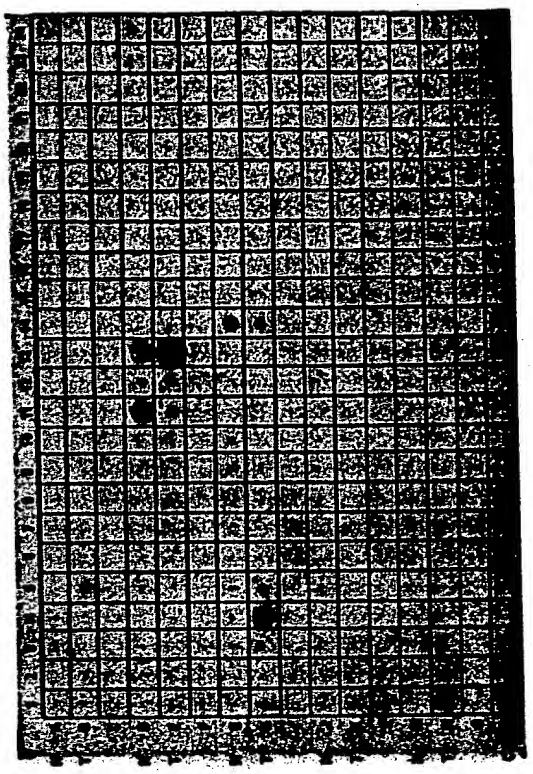


Fig. 1. Dot blot analysis of *PAGE4* probe. Hybridization of *PAGE4* probe on a matched tumor (*T*)/normal (*N*) expression array membrane (Clontech) showed specific hybridization of *PAGE4* in different samples. Samples on the blot are from kidney (*A* and *B*, *1–11*), breast (*D* and *E*, *1–9*), prostate (*D* and *E*, *11–13*), uterus (*G* and *H*, *1–7*), ovary (*G* and *H*, *9–12*), cervix (*G* and *H*, *14*), colon (*J* and *K*, *1–11*), lung (*J* and *K*, *13–15*), stomach (*M* and *N*, *1–8*), rectum (*M* and *N*, *10–16*), and small intestine (*M* and *N*, *19*).

***PAGE4* mRNA Is Expressed in Epithelial Cells of Normal Prostate and Prostate Cancer.** The RNA used for both dot blot and RT-PCR analyses was extracted from whole tissue, which consisted of mixed populations of epithelial cells, smooth muscle cells, and fibroblasts, as well as other cell types. To determine the cell types that express the *PAGE4* mRNA in normal prostate as well as in prostate cancer tissue, we used *in situ* hybridization with biotin-labeled *PAGE4* cDNA as described in "Materials and Methods." As shown in Fig. 2, *PAGE4* mRNA was highly expressed in prostatic epithelial cells of normal prostate (*A*) and in prostate cancer tissue (*B*). There was no detectable signal in cells of the stromal compartment, which suggests that *PAGE4* was specifically expressed in the epithelial cells of the prostate.

***PAGE4* mRNA Encodes for a M_r 16,000 Cytoplasmic Protein.** We reported previously that the *PAGE4* transcript has a predicted open reading frame of 102 amino acids,

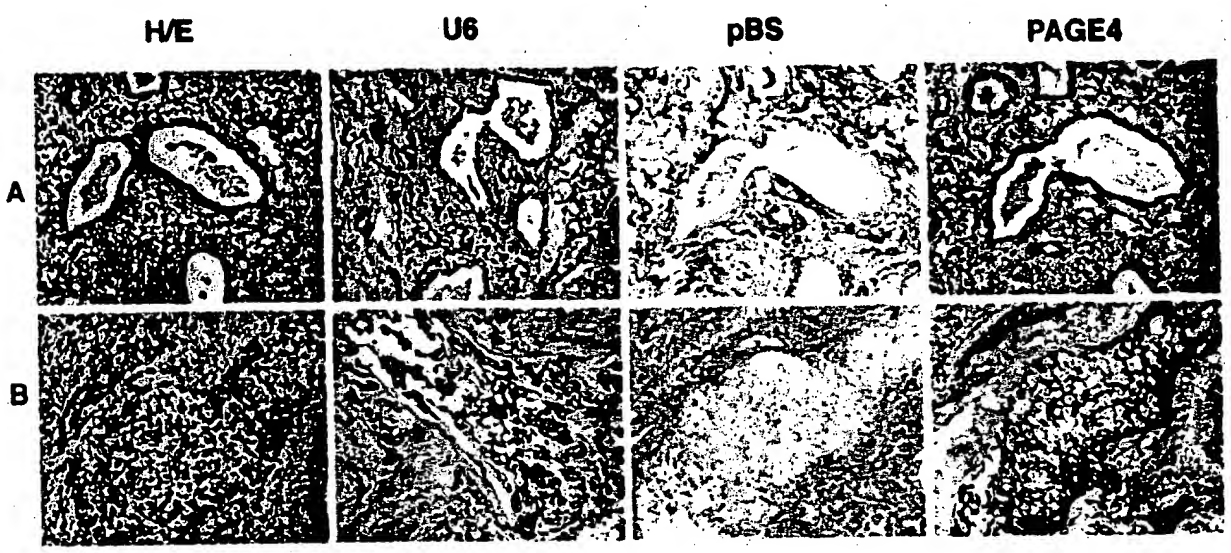


Fig. 2. In situ localization of *PAGE4* mRNA expression. *In situ* hybridization of normal prostate and prostate cancer tissue. *A*, H&E (*H/E*) stains of a representative field of benign prostatic ducts. *PAGE4* is strongly expressed in atrophic basaloid type epithelium. *B*, an example of solid lobular adenocarcinoma, Gleason grade 4 (score 8/10) strongly expressing *PAGE4*. The *PAGE4* signal is comparable in intensity with U6 positive control, in contrast to pBS negative control.

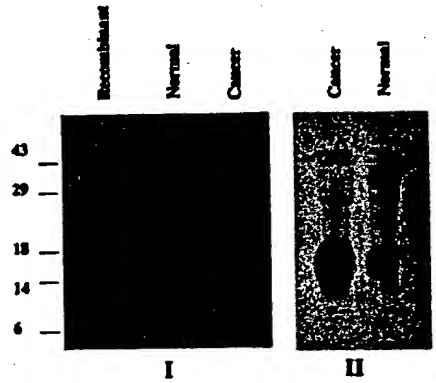


Fig. 3. Western blot analysis of PAGE4 protein. PAGE4 protein is detected in normal prostate and prostate cancer lysate. Protein extracts from normal prostate and prostate cancer (80 μg each) were analyzed on a Tris-tricine 16.5% PAGE and blotted with either antipeptide (*I*) or anti-full-length PAGE4 (*II*) polyclonal antibodies. Purified recombinant PAGE4 (10 ng) protein was used as a positive control.

which can generate a protein of M_r 11,900 in size. To determine the size of PAGE4 protein expressed in normal prostate and in prostate cancer tissue, we generated polyclonal rabbit antibodies against a chemically synthesized PAGE4 peptide as well as a recombinant PAGE4 protein as described in the "Materials and Methods" section. As shown in Fig. 3, a specific band at a M_r of ~16,000 was detected in protein extracts from normal prostate and prostate cancer by both antibodies. The M_r 16,000 kDa band comigrated with the recombinant PAGE4 protein, which indicated that the band detected in normal prostate and in prostate cancer was PAGE4. We did not detect this band using IgG prepared from preimmune rabbit serum or using a brain tissue extract but we did find expression of PAGE4 in placenta which is known to express *PAGE4* mRNA (data not shown).

To determine the location of PAGE4 protein in the cell, we performed two sets of experiments. In the first experiment, we prepared nuclear, cytoplasmic, and membrane fractions from NIH3T3 and PC3 cells stably expressing *PAGE4*. As shown in Fig. 4*A*, a specific M_r 16,000 PAGE4 product, which comigrates with recombinant PAGE4, was detected only in the cytoplasmic fraction of both of the cell lines, and not in the nuclear or membrane fractions. In a second experiment, we transfected 293T cells and infected NIH3T3 cells transiently with an expression (pCI-*PAGE4*) plasmid and performed an immunofluorescence experiment with anti-PAGE4 antibody. As shown in Fig. 4*B*, an intense cytoplasmic staining was observed in both NIH3T3 and 293T cells, which were transfected with *PAGE4* cDNA, indicating that PAGE4 protein is localized in the cytoplasm of the cells. No staining was observed in cells transfected with empty vector (data not shown).

PAGE4-regulated Gene Expression in Infected NIH3T3 and PC3 Cell Lines. Although many CT antigens have now been reported, very little is known about their biological functions. We investigated the ability of *PAGE4* to alter gene expression in the mouse NIH3T3 cell line. NIH3T3 cells do not express *PAGE4* mRNA. We generated several stable cell lines expressing the *PAGE4* transcript and protein; for the analysis, we pooled the RNA from three different stable of lines NIH3T3. To identify the transcript, which could be differentially regulated in the presence or absence of *PAGE4*, we used cDNA microarray hybridization, because it allowed us to screen numerous transcripts simultaneously. The mouse array used in this study contains 2688 cDNAs including known genes and ESTs from unknown genes (Mm-OC-6.1p).[4] Genes that were either up- or down-regulated at least

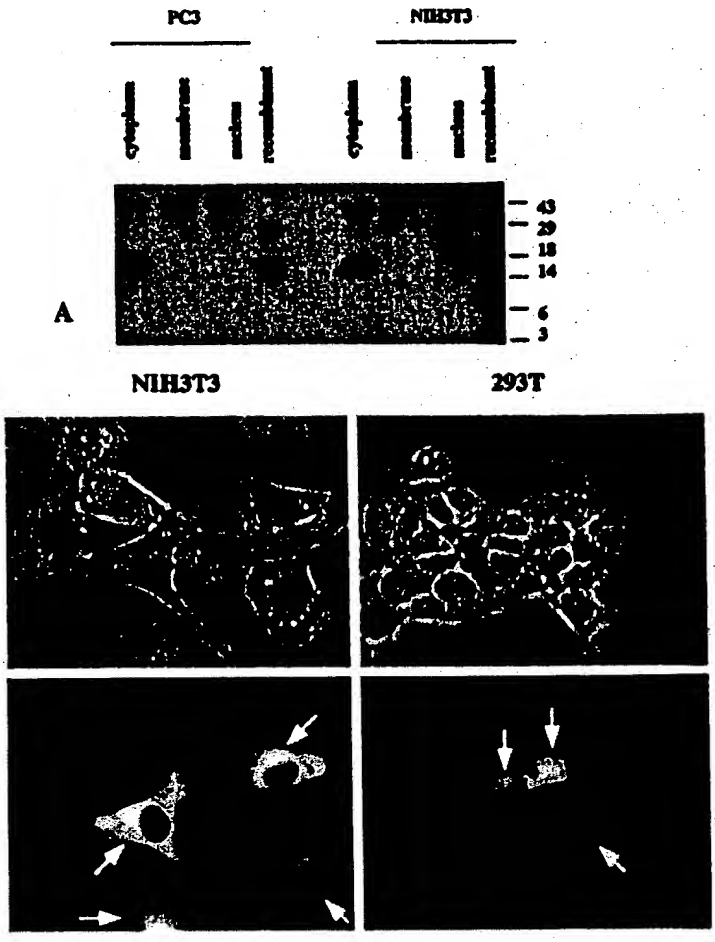


Fig. 4. Detection of PAGE4 protein on transfected cells. *A*, PAGE4 protein is detected in the cytoplasmic fraction of the cell. Protein extracts (cytoplasm, membrane, and nuclear fraction) from *PAGE4*-transfected NIH3T3 and PC3 cell lines were analyzed on a Tris-tricine 16.5% PAGE. Ten μg of protein extract was loaded on each lane. For positive control, 10 ng of recombinant *PAGE4* was loaded. *B*, *PAGE4*-transfected NIH3T3 and 293T cells were fixed, permeabilized using saponin, and incubated with anti-PAGE4 antibodies, followed by an antirabbit IgG-rhodamine conjugate. Intense cytoplasmic staining (*arrows*) was observed in both cell lines.

Table 1 Genes up/down-regulated in PAGE4 and β galactosidase (β Gal) stable cell lines

Gene	NIH3T3		PC3	
	PAGE4	β Gal	PAGE4	β Gal
Dlk1	↑[a]	↑	noc	noc
IGFBP 2	↑	↑	–	–
LpL	↓	–	–	–
SCGN-10[b]	noc	noc	↓	↓

[a] ↑, up-regulated; ↓, down-regulated; –, no change; noc, not on chip.
[b] SCGN-10, uperior cervical ganglia, neural-special 10.

2.5-fold in *PAGE4*-transfected cells, but not in β-galactosidase-transfected cells, were considered for analysis.

Only three genes were found to be altered by 2.5-fold or more (Table 1). One of the genes, that was specifically modulated in *PAGE4*-transfected cells as compared with vector-alone-transfected cells, was *LPL*. LPL is the primary enzyme responsible for the conversion of lipoprotein triglycerides into free fatty acids and monoglycerides. *LPL* was reproducibly down-regulated in *PAGE4*-expressing stable cells (Fig. 5*A*), as compared with vector-alone-transfected or β-galactosidase-transfected cell lines. Two other genes, *Dlk1* and *IGFbp-2*, were up-regulated in *PAGE4*-transfected cells, but they were also up-regulated in β-galactosidase-transfected NIH3T3 cells (Table 1). Thus, this effect of *PAGE4* was not specific. Down-regulation of *LPL* expression in *PAGE4* stable lines was verified by Northern blot analysis. RNA was isolated from each of three independent cell lines that were transfected with either *PAGE4* or empty vector, and that were subjected to Northern analysis using a radiolabeled *LPL* cDNA probe. As shown in Fig. 5*B*, all three of the cell lines stably expressing *PAGE4* had undetectable *LPL* mRNA expression as compared with the cells infected with empty vector or vector expressing the *β-galactosidase* gene.

We then investigated the ability of *PAGE4* to alter gene expression in the PC3 prostate cancer cell line. PC3 is an androgen-independent prostate-cancer cell line and does not express *PAGE4* mRNA.[5] We generated several stable cell lines expressing the PAGE4 transcript and protein, and we analyzed RNA expression with an array that contains 6538 human cDNAs including known genes and ESTs from unknown genes.[4] We found that none of the cDNAs was significantly modulated in the PC3 cells expressing *PAGE4*

[5] Iavarone and Bera, unpublished observations.

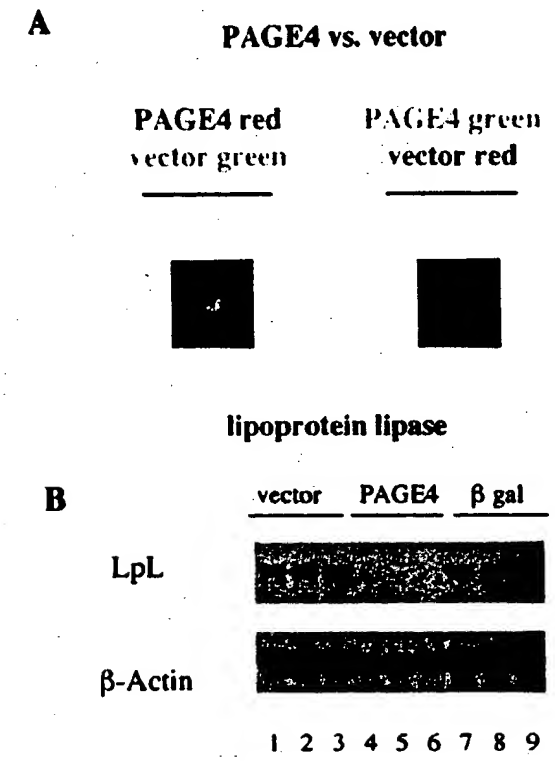


Fig. 5. PAGE4-mediated differential gene expression in transfected NIH3T3 cells. *A*, gene expression profiling of *PAGE4*-transfected and vector-alone-transfected NIH3T3 cells by microarray analysis. The same spot (*LPL*) is detected by two independent experiments in which *PAGE4*-transfected cell RNA is labeled with red or green dye. In *B*, LPL transcript is down-regulated in *PAGE4*-transfected 3T3 cells. An equal amount (2 μg) of polyadenylate RNA from NIH3T3 cells (three independent clones), transfected with either *PAGE4* or vector alone, was size-fractionated and transferred to nitrocellulose filter. Filters were then hybridized with labeled LPL cDNA as described in "Materials and Methods." The filters were stripped and reprobed with labeled actin to determine equal loading of RNA in the membrane.

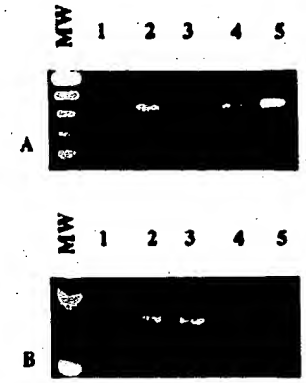


Fig. 6. RT-PCR and PCR analyses of *LPL* gene in PC3 cell line. Ethidium bromide-stained 1.2% agarose gels: in *A*, *LPL* cDNA was amplified with specific primers for the 3' untranslated region of *LPL* mRNA. A specific band was amplified by using normal prostate cDNA (Clontech; *Lane 2*), PC3 genomic DNA (*Lane 4*) and human genomic DNA (*Lane 5*). No band was amplified without template (*Lane 1*) and using PC3 cDNA (*Lane 3*). In *B*, integrity of each mRNA and genomic DNA samples used was tested by amplifying the *actin* gene using specific primers for actin.

including LPL, which was down-regulated in mouse array analysis (see below).

Discussion

PAGE4 is a CT antigen expressed in normal prostate, testis, and placenta, and highly expressed in prostate and uterine cancer. In this report, we demonstrate that *PAGE4* mRNA expression is restricted to the epithelial cells of normal prostate and prostate carcinomas. The PAGE4 protein is M_r 16,000 in size and is localized in the cytoplasmic compartment of the cell. To investigate the function of *PAGE4*, we introduced *PAGE4* into a mouse fibroblast cell line (NIH3T3) and also into a prostate cancer cell line (PC3) and found that LPL is down-regulated in NIH3T3 cells expressing *PAGE4*, but not in PC3 cells.

Expression of PAGE4 Affects *LPL* Gene Expression. LPL, a rate-limiting enzyme responsible for the hydrolysis of circulating triglyceride, is bound to the luminal surface of capillary endothelium in adipose tissue and muscle (10). Although LPL is functional at the surface of the endothelial cells, it is not clear which cells synthesize the enzyme and what the cellular distribution of LPL is within the tissue. It has been reported that the expression of the *LPL* gene is regulated by progesterone and protein kinase A in undifferentiated hepatoma cells (11). Reduction in LPL activity has been implicated in cachexia in cancer patients. Cachexia is a common feature of malignant disease and is characterized by marked weight loss, anorexia, and extensive breakdown of body fat and skeletal protein because of homeostatic disturbances (12). It has been reported that the *LPL* gene in humans is localized on chromosome 8p22 (13), and there is a frequent allelic loss of chromosome 8p22 loci in human prostate cancer. Because we observed down-regulation of LPL in NIH3T3 cells expressing *PAGE4*, but not in the prostate cancer cell line PC3, we investigated whether the *LPL* gene is expressed in the PC3 cell line. As shown in Fig. 6*A*, *Lane 3*, the *LPL* gene is not expressed in the PC3 cell line, thus, *PAGE4* could not possibly down-regulate LPL expression in this cell line. To determine whether the *LPL* gene is deleted in PC3 cells, we did a PCR analysis on chromosomal DNA isolated from PC3 cells. As shown in Fig. 6*A*, *Lane 4*, the intensity of the PCR product from PC3 DNA is one-half the intensity of the PCR product (Fig. 6*A*, *Lane 5*) obtained from normal human chromosomal DNA, which suggests that at least one of the *LPL* alleles is deleted in PC3 cells. Further analysis of the PC3 cell line is needed to establish the exact copy number of the *LPL* gene in PC3 cells, and the significance of the LPL deletion in prostate cancer and its down-regulation by *PAGE4* gene needs further study.

Although CT antigens have been identified for over a decade, no function has been described for these antigens in the literature. These antigens are encoded quite frequently by genes located on the X chromosome. There is speculation that some CT gene products are transcriptional factors, but there is no direct experimental evidence supporting that concept. Two of the previously reported CT antigens, SCP-1 and CT9 have small basic domains and several conserved motifs, which are characteristics of DNA-binding proteins

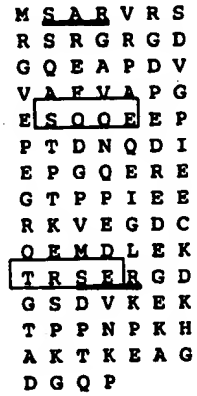


Fig. 7. Amino acid analysis of PAGE4. Amino acid sequence analysis of PAGE4 was performed using the PROSITE program. *Underlined*, two potential protein kinase C phosphorylation sites (amino acids 2–4 and 73–75) contained in PAGE4 protein. *Boxed*, two potential casein kinase 2 phosphorylation sites (amino acids 30–33 and 71–74).

(14, 15). Amino acid sequence analysis of *PAGE4* showed no such motifs in the PAGE4 protein.

At this point, it is not clear whether the regulatory effect of *PAGE4* on the *LPL* gene is direct or indirect. Because PAGE4 protein is localized in the cytoplasm of the cell, it is more likely that *PAGE4* indirectly regulates the expression of the *LPL* gene. It is also possible that *PAGE4* is posttranslationally modified and involved in a signal transduction pathway. Analysis of the *PAGE4* amino acid sequence using the PROSITE program of the Swiss Institute of Bioinformatics ExPASy proteomics server[6] (16, 17) predicts several potential phosphorylation sites. As shown in Fig. 7 there are two casein kinase 2 (SQQE and TRSE) sites and two protein kinase C phosphorylation sites (SAR and SER). The significance of these phosphorylation sites, as well as the mechanism by which *PAGE4* alters the expression of the *LPL* gene, is yet to be determined. To our knowledge, this is the first report of CT antigen-mediated gene regulation in the literature.

PAGE4, a CT Antigen, Is a Potential Target for Cancer Vaccines. CT antigens are a distinct class of antigens that have a restricted pattern of expression in essential tissue and aberrant expression in many different tumor types (18). CT antigens are often expressed at higher levels in testis and placenta, which are known to express only low amounts of MHC class I molecules. Thus, expression of CT antigens in these normal tissues should not lead to T-cell activation and this makes these antigens attractive candidates for cancer vaccines (19). Our earlier studies demonstrated that *PAGE4* is expressed in prostate and uterine cancer. Data presented here indicate that *PAGE4* is also expressed in cervical, ovarian, and kidney cancer and increases its attractiveness as a cancer vaccine target.

Acknowledgments

We thank Verity Fogg for cell culture and Anna Mazzuca for editorial assistance.

[6] Internet address: http://www.expasy.ch.

References

1. Vasmatzis, G., Essand, M., Brinkmann, U., Lee, B., and Pastan, I. Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. Proc. Natl. Acad. Sci. USA, *95:* 300–304, 1998.

2. Brinkmann, U., Vasmatzis, G., Lee, B., Yerushalmi, N., Essand, M., and Pastan, I. *PAGE-1*, an X chromosome-linked GAGE-like gene that is expressed in normal and neoplastic prostate, testis, and uterus. Proc. Natl. Acad. Sci. USA, *95:* 10757–10762, 1998.

3. Kinisella, T. M., and Nolan, G. P. Episomal vectors rapidly and stably produce high-titer recombinant retrovirus. Hum. Gene Ther., *7:* 1405–1413, 1996.

4. Liu, X. F., Olsson, P., Wolfgang, C. D., Bera, T. K., Duray, P., Lee, B., and Pastan, I. PRAC: a novel small nuclear protein that is specifically expressed in human prostate and colon. Prostate, *47:* 125–131, 2001.

5. Kumar, V., and Collins, F. H. A technique for nucleic acid *in situ* hybridization to polytene chromosomes of mosquitoes in the *Anopheles gambiae* complex. Insect Mol. Biol., *3:* 41–47, 1994.

6. Wolfgang, C. D., Essand, M., Vincent, J. J., Lee, B., and Pastan, I. TARP: a nuclear protein expressed in prostate and breast cancer cells derived from an alternate reading frame of the T cell receptor γ chain locus. Proc. Natl. Acad. Sci. USA, *97:* 9437–9442, 2000.

7. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res., *11:* 1475–1489, 1983.

8. Sladek, F., M., Zhong, W. M., Lai, E., and Darnell, J. E., Jr. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev., *4:* 2353–2365, 1990.

9. Chang, K., Pastan, I., and Willingham, M. C. Isolation and characterization of a monoclonal antibody, K1, reactive with ovarian cancers and normal mesothelium. Int. J. Cancer, *50:* 373–381, 1992.

10. Goldberg, I. J., and Merkel, M. Lipoprotein lipase: physiology, biochemistry, and molecular biology. Front. Biosci., *6:* 388–405, 2001.

11. Peinadoonsurbe, J., Staels, B., Deeb, S., and Auwerx, J. Lipoprotein-lipase expression in undifferentiated hepatoma-cells is regulated by progesterone and protein kinase-A. Biochemistry, *31:* 10121–10128, 1992.

12. DeWys, W. D., Begg, C., Lavin, P. T., Band, P. R., Bennett, J. M., Bertino, J. R., Cohen, M. H., Douglass, H. O., Engstrom, P. F., Ezdinli, E. Z., Horton, J., Johnson, G. J., Moertel, C. G., Oken, M. M., Perlia, C., Rosenbaum, C., Silverstein, M. N., Skeel, R. T., Sponzo, R. W., and Tormey, D. C. Prognostic effect of weight-loss prior to chemotherapy in cancer-patients. Am. J. Med., *69:* 491–497, 1980.

13. Bova, G. S., Carter, B. S., Bussemakers, M. J., Emi, M., Fujiwara, Y., Kyprianou, N., Jacobs, S. C., Robinson, J. C., Epstein, J. I., Walsh, P. C., and Isaacs, W. B. Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer. Cancer Res., *53:* 3869–3873, 1993.

14. Jones, M. H., Numata, M., and Shimane, M. Identification and characterization of *BRDT*: a testis-specific gene related to the bromodomain genes *RING3* and *Drosophila fsh*. Genomics, *45:* 529–534, 1997.

15. Meuwissen, R. L. J., Offenberg, H. H., Dietrich, A. J. J., Riesewijk, A., Vaniersel, M., and Heyting, C. A coiled-coil related protein-specific for synapsed regions of meiotic prophase chromosomes. EMBO J., *11:* 5091–5100, 1992.

16. Appel, R. D., Bairoch, A., and Hochstrasser, D. F. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem. Sci., *19:* 258–260, 1994.

17. Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. The PROSITE database, its status in 1999. Nucleic Acids Res., *27:* 215–219, 1999.

18. Ono, T., Kurashige, T., Harada, N., Noguchi, Y., Saika, T., Niikawa, N., Aoe, M., Nakamura, S., Higashi, T., Hiraki, A., Wada, H., Kumon, H., Old, L. J., and Nakayama, E. Identification of proacrosin binding protein sp32 precursor as a human cancer/testis antigen. Proc. Natl. Acad. Sci. USA, *98:* 3282–3287, 2001.

19. Moingeon, P. Cancer vaccines. Vaccine, *19:* 1305–1326, 2001.

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman‖, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. ***Molecular & Cellular Proteomics 1:37–45, 2002.***

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1*, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)[1] have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q−, 11p−, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

From the ‡Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and **Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ¶AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ‖UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

[1] The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

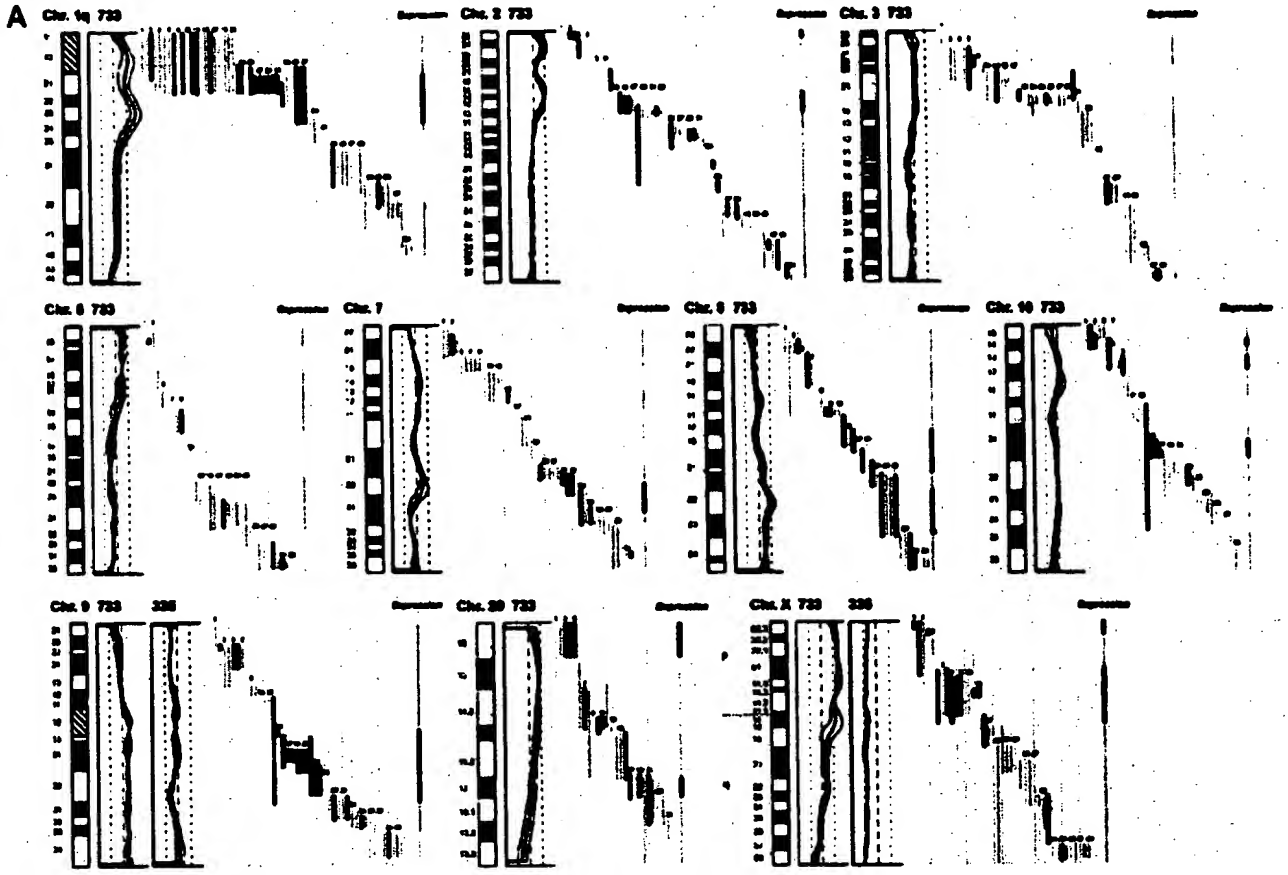


FIG. 1. **DNA copy number and mRNA expression level.** Shown from *left* to *right* are chromosome (*Chr.*), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. *A*, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. *B*, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (*left*). The *bold curve* in the ratio profile represents a mean of four chromosomes and is surrounded by *thin curves* indicating one standard deviation. The *central vertical line* (*broken*) indicates a ratio value of 1 (no change), and the *vertical lines* next to it (*dotted*) indicate a ratio of 0.5 (*left*) and 2.0 (*right*). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the *right* of the invasive tumor profile. The *colored bars* represents one gene each, identified by the running *numbers above* the *bars* (the name of the gene can be seen at www.MDL.DK/sdata.html). The *bars* indicate the purported location of the gene, and the *colors* indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (*black*), >2-fold decrease (*blue*), no significant change (*orange*). The *bar* to the *far right*, entitled *Expression* shows the resulting change in expression along the chromosome; the *colors* indicate that at least half of the genes were up-regulated (*black*), at least half of the genes down-regulated (*blue*), or more than half of the genes are unchanged (*orange*). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at −80 °C. Total RNA was isolated using the RNAzol B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)$^+$ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript® choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the MEGAscrip® *in vitro* transcription kit (Ambion). Biotin-labeled CTP and UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94 °C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6× SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95 °C for 5 min, subsequently cooled to 40 °C, and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40 °C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6× SSPE-T at 25 °C followed by 4 washes in 0.5× SSPE-T at 50 °C. The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 μg/ml (Molecular Probes) in 6× SSPE-T

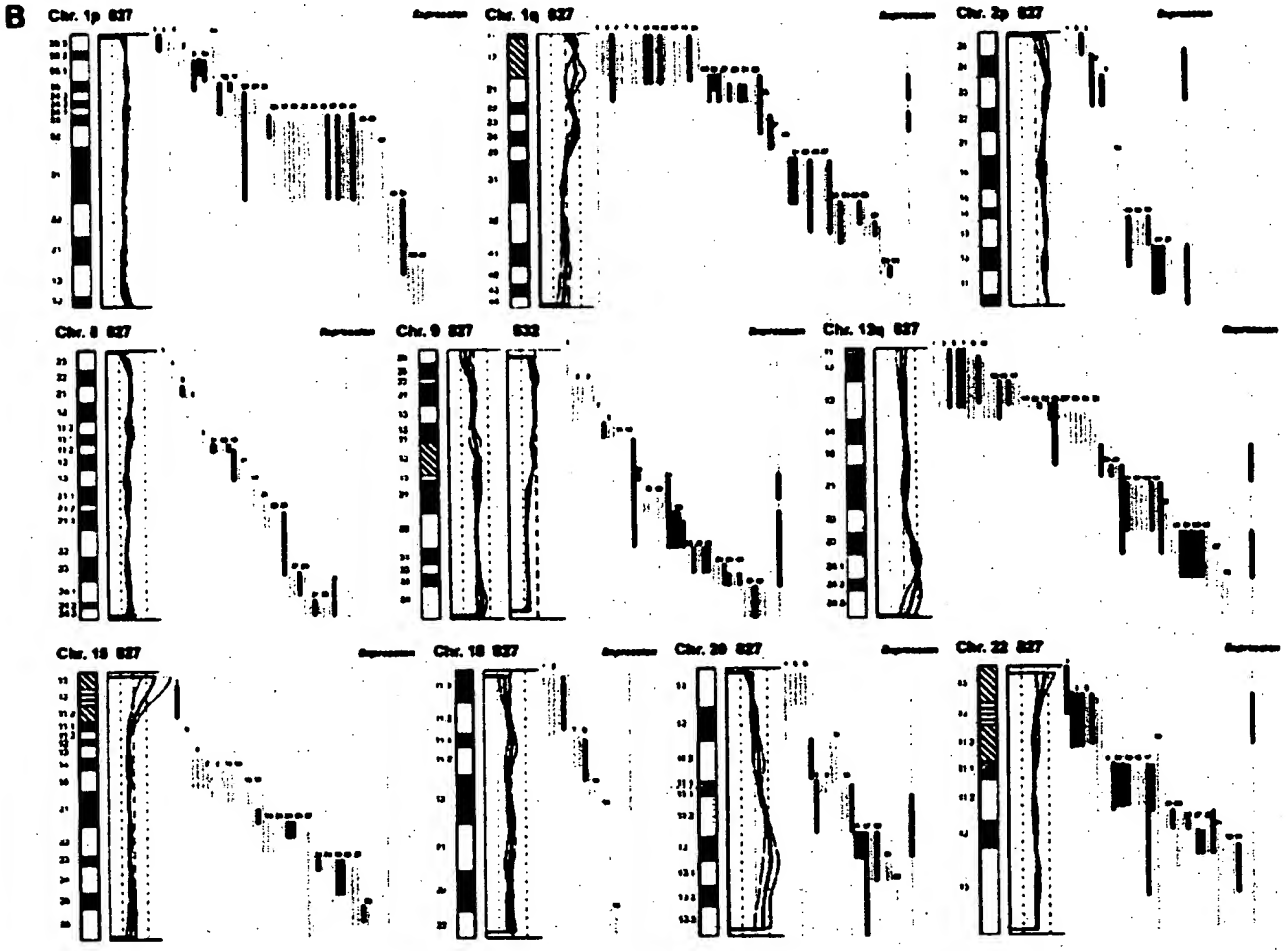


FIG. 1—*continued*

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μl for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at −20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/celis.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-labeled reference DNA (200 ng), and human Cot-1 DNA (20 μg) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μg/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

TABLE I

Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (If CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (If expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%

Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p−, 9q22-q33−, and X−, and 7+, 9q−, and Y−, respectively. Both invasive tumors showed changes (1q22-24+, 2q14.1-qter−, 3q12-q13.3−, 6q12-q22−, 9q34+, 11q12-q13+, 17+, and 20q11.2-q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1*A*) and 20q12 in TCC 827 (Fig. 1*B*).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21-q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, *top*). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, *top*). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, *bottom*). Furthermore, as a control we looked at areas with no alter-

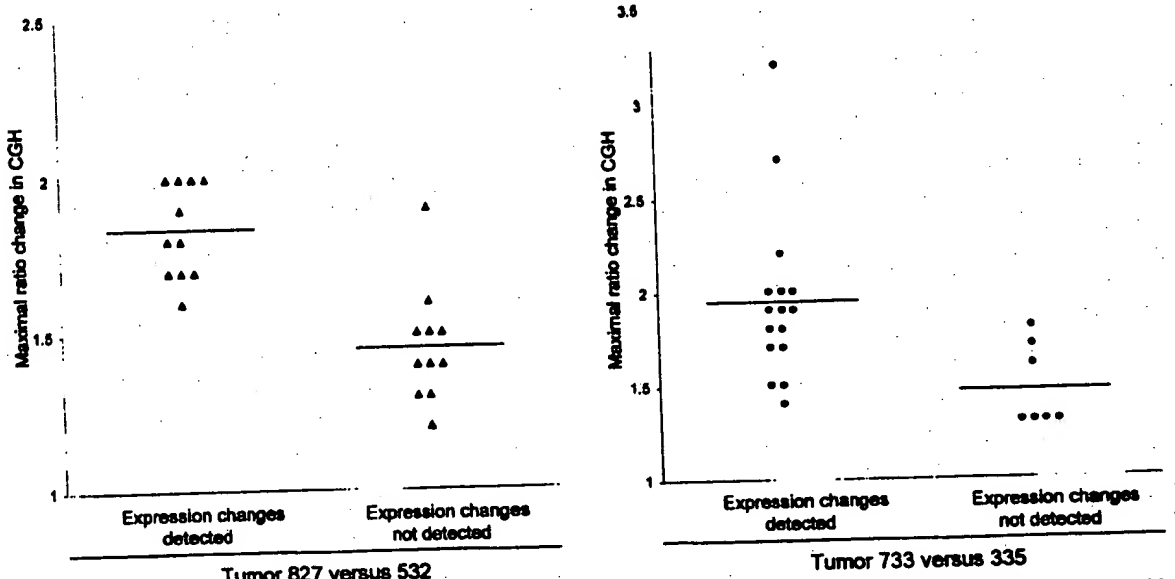


FIG. 2. **Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring.** The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (♦) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the *right* in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1*A*). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

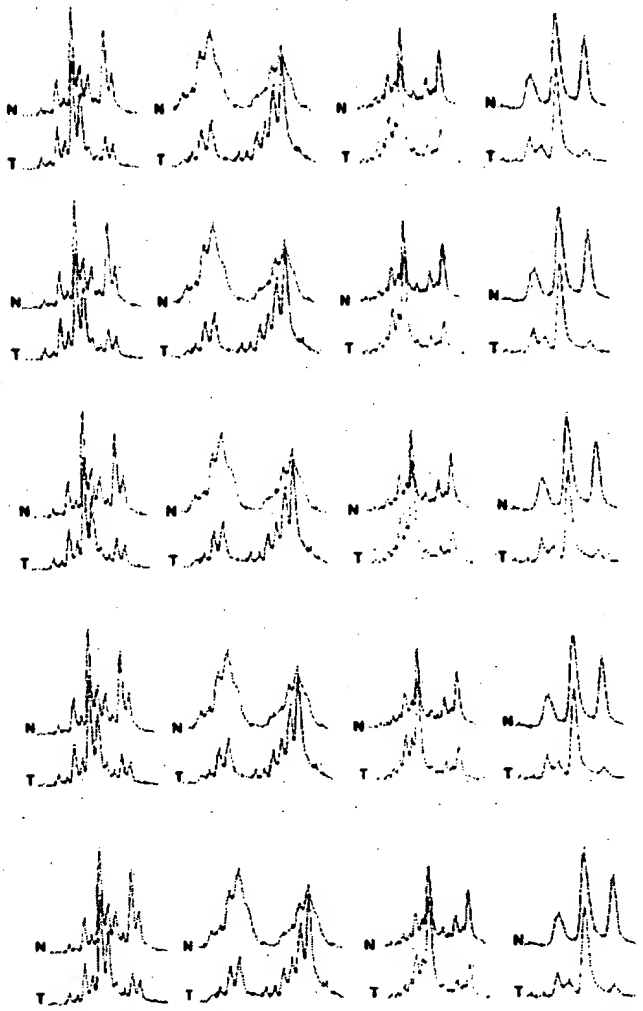


FIG. 3. **Microsatellite analysis of loss of heterozygosity.** Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (*a*) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (*b*) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (*c*) at chromosome 2p23 by D2S2251 close to general β-spectrin (gene number 11 on Fig. 1) and of (*d*) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The *upper curves* show the electropherogram obtained from normal DNA from leukocytes (*N*), and the *lower curves* show the electropherogram from tumor DNA (*T*). In all cases one allele is partially lost in the tumor amplicon.

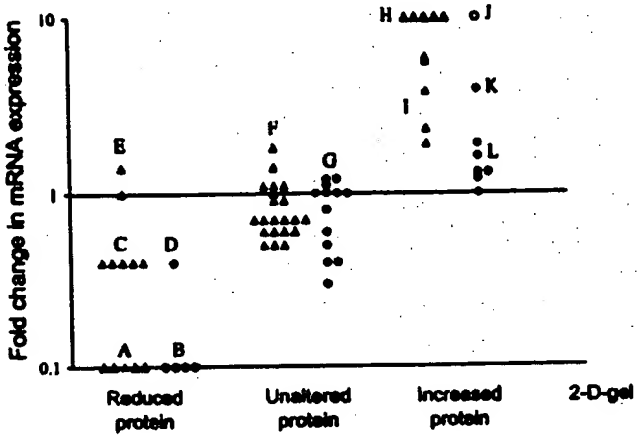


FIG. 4. **Correlation between protein levels as judged by 2D-PAGE and transcript ratio.** For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (*horizontal axis*). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (*vertical axis*). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; △, mRNAs that were scored as absent in the invasive tumors (along *horizontal axis*) or as absent in non-invasive reference (*top* of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲△) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group *A* (from *left*), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytokeratin 15, and cytokeratin 13; *B* (from *left*), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytokeratin 13, and calcyclin; *C* (from *left*), α-enolase, hnRNP B1, 28-kDa heat shock protein, 14–3-3-ε, and pre-mRNA splicing factor; *D*, mesothelial keratin K7 (type II); *E* (from *top*), glutathione *S*-transferase-π and mesothelial keratin K7 (type II); *F* (from *top* and *left*), adenylyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ-actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β-1 subunit; *G*, (from *top* and *left*), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytokeratin 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione *S*-transferase-π, and keratin 8; *H* (from *left*), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15 hydroxyprostaglandin dehydrogenase; *I* (from *top*), prolyl 4-hydroxylase β-subunit, cytokeratin 20, cytokeratin 17, prohibition, and fructose 1,6-biphosphatase; *J* annexin II; *K*, annexin IV; *L* (from *top* and *left*), 90-kDa heat shock protein, prolyl 4-hydroxylase β-subunit, α-enolase, GRP 78, cyclophilin, and cofilin.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

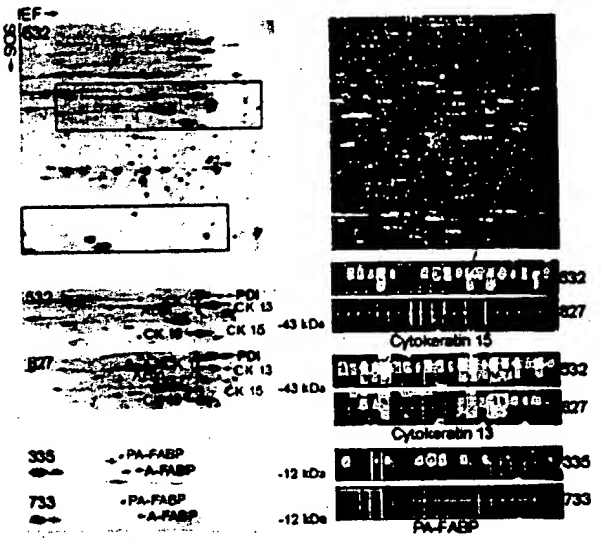


FIG. 5. **Comparison of protein and transcript levels in invasive and non-invasive TCCs.** The *upper part* of the figure shows a 2D gel (*left*) and the oligonucleotide array (*right*) of TCC 532. The *red rectangles* on the *upper gel* highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (*red annotation*). The tile on the array containing probes for cytokeratin 15 is enlarged *below* the array (*red arrow*) from TCC 532 and is compared with TCC 827. The *upper row* of *squares* in each tile corresponds to perfect match probes; the *lower row* corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as *black*, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the *bottom* of the figure (*left*) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the *right* we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). *IEF*, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a know chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II

Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration[a]	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres[a]	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up[b]	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

[a] Abs, absent; Pres, present.
[b] In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (*e.g.* in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p−, 9q−, 1q+, Y− (2, 6), and in pT1 tumors, 2q−,11p−, 11q−, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p− and 9q22-q33− and 9q− and Y−, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p−, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicate that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (*e.g.* signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Vikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

REFERENCES

1. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* **17**, 643–649
2. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46
3. de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* **83**, 157–161
4. Brungier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* **12**, 1747–1753
5. Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* **50**, 1459–1463
6. Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) c-*myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am J. Pathol.* **146**, 1131–1139
7. Richter, J., Jiang, F., Gorog, J. P., Sartorious, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* **57**, 2860–2864
8. Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* **153**, 1615–1621
9. Bruch, J., Wohr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* **23**, 167–174
10. Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* **15**, 3555–3560
11. Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* **185**, 345–351
12. Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* **110**, 87–93
13. Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367
14. Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* **59**, 167–177
15. Celis, J. E., Ostergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* **56**, 4782–4790
16. Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL
17. Ohlsson, R., Tycko, B., and Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* **14**, 435–438
18. Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121
19. Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170
20. Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpa, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* **59**, 3889–3892
21. Cui, H., Hedborg, F., He, L., Nordenskjold, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* **57**, 4469–4473
22. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* **285**, 251–254
23. Tsao, J., Yatabe, Y., Marki, I. D., Hajyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* **29**, 26–32
24. Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10632–10636
25. Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
26. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934

Scheme for Ranking Potential HLA-A2 Binding Peptides Based on Independent Binding of Individual Peptide Side-Chains

Kenneth C. Parker,[1]* Maria A. Bednarek,† and John E. Coligan*

*Biological Resources Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; and †Merck Sharp and Dohme Research Laboratories, Rahway, NJ 07065

ABSTRACT. A method to predict the relative binding strengths of all possible nonapeptides to the MHC class I molecule HLA-A2 has been developed based on experimental peptide binding data. These data indicate that, for most peptides, each side-chain of the peptide contributes a certain amount to the stability of the HLA-A2 complex that is independent of the sequence of the peptide. To quantify these contributions, the binding data from a set of 154 peptides were combined together to generate a table containing 180 coefficients (20 amino acids × 9 positions), each of which represents the contribution of one particular amino acid residue at a specified position within the peptide to binding to HLA-A2. Eighty peptides formed stable HLA-A2 complexes, as assessed by measuring the rate of dissociation of β_2m. The remaining 74 peptides formed complexes that had a half-life of β_2m dissociation of less than 5 min at 37°C, or did not bind to HLA-A2, and were included because they could be used to constrain the values of some of the coefficients. The "theoretical" binding stability (calculated by multiplying together the corresponding coefficients) matched the experimental binding stability to within a factor of 5. The coefficients were then used to calculate the theoretical binding stability for all the previously identified self or antigenic nonamer peptides known to bind to HLA-A2. The binding stability for all other nonamer peptides that could be generated from the proteins from which these peptides were derived was also predicted. In every case, the previously described HLA-A2 binding peptides were ranked in the top 2% of all possible nonamers for each source protein. Therefore, most biologically relevant nonamer peptides should be identifiable using the table of coefficients. We conclude that the side-chains of most nonamer peptides to the first approximation bind independently of one another to the HLA-A2 molecule. *Journal of Immunology,* 1994, 152: 163.

MHC class I molecules are normally expressed on the cell surface in a stable complex, with any one of a large number of peptides generated upon proteolysis of intracellular proteins (1, 2). In theory, each allelic variant of a class I MHC molecule selects these peptides based on the complementary structure of the peptide and the polymorphic pockets within the peptide-binding groove (3, 4). In the past, motifs specific to individual class I molecules have been determined by comparing the sequences of endogenous peptides isolated from purified class I molecules (5, 6), or by comparing the sequences of peptides that are known to bind to each class I molecule (7). In every case studied so far, at certain positions within the peptide, one aa[2] or a small number of related aa are found to be nearly invariant; these aa are called dominant anchor residues (5) and appear to "anchor" the peptide into the class I peptide binding site by having a structure that is complementary to a pocket of the peptide-binding groove. For example, endogenous peptides isolated from purified HLA-A2 contain as dominant anchor residues Leu or Met at P2, and Val or Leu at P9 (5, 8), which are thought to bind in the B and F pockets, respectively (4). Some of the other positions within the endogenous peptides are also enriched for specific aa; these are defined as auxiliary anchor residues (5). In many cases, it is not clear to what degree the allele-specific peptide-binding motifs consisting of dominant and auxiliary anchor residues are a consequence of the requirements

Received for publication June 25, 1993. Accepted for publication October 6, 1993.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Address correspondence and reprint requests to Dr. Kenneth C. Parker, Building 4, Room 413, BRB, NIAID, NIH, Bethesda, MD 20892.

[2] Abbreviations used in this paper: aa, amino acid residue; GF, gel filtration; IBS, independent binding of side-chains; M1, influenza A matrix protein peptide 58–66 (GILGFVFTL); P1, position 1 in a peptide; PΩ, last position in a peptide; β_2m, β_2-microglobulin.

0022-1767/94/$02.00

for peptide binding (9), or whether part of the motif is a consequence of Ag processing (5). These motifs have occasionally proven useful for localizing the optimal class I binding peptide within the sequence of a longer peptide that is known to contain a HLA-A2-restricted T cell epitope (10, 11). However, in other cases, no obvious motif is present within the antigenic peptide (12–15), possibly reflecting the limitations in our knowledge about what are acceptable variations in peptide-binding motifs.

To determine to what degree the dominant anchor residues or the auxiliary anchor residues of the HLA-A2 motif are important for peptide binding, we have extended our previous peptide-binding study (16) to include a more extensive set of nonapeptides, many of which are interrelated by single aa substitutions. These new data indicate that at each position within the peptide, some aa are more favorable than others, regardless of the sequence of the rest of the peptide; therefore, it should theoretically be possible to improve upon predictions based solely on the anchor residues at P2 and P9 that are specific to HLA-A2 binding peptides. We present a table of 180 coefficients specific for each of the 20 aa at each of the 9 positions within the peptide. This table of coefficients incorporates all of the data that we have collected and can be used to predict the stability of HLA-A2 complexes containing any desired peptide. A mathematical combinatorial approach similar to that used to generate the HLA-A2 binding coefficients could also be applied to other macromolecular interactions, such as between oligonucleotides and DNA-binding proteins.

Materials and Methods

Peptides

Peptides were synthesized and purified as described (17).

HLA-A2 binding assays

Native isoelectric focusing gel and GF peptide binding assays were used as described (17). These assays measure peptide binding indirectly by monitoring the ability of peptides to promote incorporation of ^{125}I-labeled β_2m into HLA-A2/β_2m/peptide heterotrimeric complexes. Because the HLA H chain is refolded from inclusion bodies prepared from *Escherichia coli*, there are no endogenous peptides present to confound the data. Instead, ^{125}I-β_2m is incorporated into HLA complexes only when an appropriate synthetic peptide is present.

Rate measurements

The stability of HLA-A2 complexes containing specific peptides was assessed by measuring the β_2m dissociation rate as described (17). Briefly, HLA complexes containing ^{125}I-β_2m were isolated by GF and β_2m dissociation was measured by means of a second round of GF. To make accurate measurements of half-lives that were less than 20 min, it was necessary to collect the purified complexes from the first round of GF directly into microcentrifuge tubes that were then maintained at 0°C. Each aliquot to be used for a time point was separately incubated from between 1 and 30 min at 37°C before the second round of GF.

Mathematical modeling

To combine the data mathematically from a large number of experiments, a Fortran program was written that could optimize for the values of all 180 (20 aa × 9 positions) coefficients with any number of simultaneous equations. The rate data for a peptide whose sequence was GILGFVFTL would be entered as follows:

$$\mathrm{ERR} = \ln(t_{1/2}) - \ln(\mathrm{G1} \times \mathrm{I2} \times \mathrm{L3} \times \mathrm{G4} \times \mathrm{F5} \times \mathrm{V6} \times \mathrm{F7} \times \mathrm{T8} \times \mathrm{L9} \times \mathrm{Constant})$$

where ERR squared equals the error function to be minimized, $t_{1/2}$ equals the measured half-life of dissociation in minutes at 37°C, G1 represents, for example, the coefficient for Gly at P1 to be determined (see Table I), and Constant equals the overall normalization constant. For the peptides that had half-lives of dissociation of <5 min, ERR was set equal to 0.0 if the second term (the logarithm of the product of the coefficients and the normalization constant) was less than $\ln(5.0)$; otherwise, $\ln(5.0)$ was subtracted from the second term. The program calculated the values for the coefficients that minimized the sum of the error functions of the 154 equations, where each equation corresponds to the rate data for one peptide. After the coefficients were calculated, the coefficients at each position were normalized by dividing by the coefficient for Ala at that same position, and all of the normalization constants for Ala at each position were combined into Constant. As a result, the coefficients for Ala are not independent, and there are only 172 independent variables: one for each aa (19 independent terms relative to Ala) at each of the 9 positions within a nonameric peptide, plus Constant. Thus, Constant equals the predicted half-time for dissociation of a complex containing the peptide AAAAAAAAA in minutes at 37°C.

In practice, to limit the number of variables to a number more appropriate for the amount of data we had collected, 82 variables were selected that appear to be the most important for determining peptide binding to HLA-A2 (see Table V). Of these variable coefficients 40 were at the dominant anchor positions at P2 and P9, which are known to be critical for peptide binding to HLA-A2 (5, 8). A total of 31 variable coefficients were selected on the basis of the relationship between the sequences of the peptides and the data that were collected. For example, a coefficient was always allowed to be a variable if the dissociation rate for an HLA-A2 complex containing a peptide was significantly different from the dissociation rate of a complex that contained a second peptide that differed in sequence only at the position of the coefficient. Eight other coefficients were allowed to be variable (D at P1; R and W at P3; W and Y at P5; W and Y at P6; and W at P7) because they correspond to aa that are chemically similar to aa that are expected to be important for binding based on the singly substituted peptide studies described above. Three coefficients at P4 (corresponding to Ile, Leu, and Phe) were allowed to be variable because peptides containing these aa formed complexes that were less stable than could otherwise be accounted for. All other coefficients were fixed at 1.0. A fortran program was written based on the algorithm of Davidon (18) to solve for all 82 variables simultaneously. A second program was written that further optimized each coefficient separately starting with the output from the first program. In addition, this program determined the maximum tolerable value for a coefficient that was present only in the set of nonbinding peptides. The coefficients (see Table V) were obtained using these two programs starting with the assumption that each coefficient (and Constant) had a value of 1.0. Almost all nonapeptides for which we have collected data were used in the calculations, and are listed (see Tables II–IV). There were two classes of exceptions. Some peptides contained such poor dominant anchor residues at P2 and P9 that including them would have no impact on the calculations. The second class of exceptions (see Table VI) were excluded because they were not consistent with the bulk of the data. Three of these peptides (see Table VIA) probably violate the requirement of independent binding of side-chains (see below). The remaining 12 peptides (see Table VIB) contained poor dominant anchor residues at P2 and P9 or poor auxiliary anchor residues at P3, and often formed complexes in relatively low yield, or highly variable yield. It is possible that the active species in some of the peptides (see Table VIB) are contaminating peptides. We are currently investigating the basis for the unusual behavior of these peptides.

Comparison of experimental to theoretical

To test how well the coefficients fit the experimental data, theoretical β_2m dissociation rates were calculated from the table of coefficients (see Table V) and the overall normalization constant. In practice, this was done using a program written using Dbase III software (Ashton-Tate,

Torrance, CA). In this program, the inputs are the table of coefficients to be used, a table containing peptide sequences (in single aa code) and experimental binding data. The program calculates the theoretical dissociation rate and the ratio of the experimental to theoretical.

Ranking of potential HLA-A2 binding peptides

Software was written in the Dbase III programming language that generates a table containing the sequence of every possible contiguous nine aa peptide starting from a protein's primary sequence. The peptides can then be indexed according to the theoretical β_2m dissociation rate, calculated using the coefficients (see Table V).

Results

β_2m dissociation rate data for pairs of peptides that differ at single aa

It has been found that the β_2m dissociation rate from HLA-A2 complexes containing peptides with a Leu at P2 and a Val at P9 varies over at least four orders of magnitude, depending on the sequence of the rest of the peptide (17). Nonetheless, the Leu at P2 in either peptide GLGGGGGGV ($t_{1/2}$ < 1 min) and in peptide LLFGYPVYV ($t_{1/2}$ 4000 min) might stabilize the corresponding HLA-A2 complex to the same degree. One way to test this idea is to compare the β_2m dissociation rates of pairs of peptides that differ in sequence by a single aa substitution. Table I contains data of this kind, listed according to the position within the peptide, and then alphabetically according to the single letter aa code of the aa to be compared (first column). For example, at the top of Table I, there are five pairs of peptides that differ only by substitution of Lys for Gly at P1. The ratio of β_2m dissociation rates obtained for HLA-A2 complexes containing each pair of peptides is shown in Table I, second column. It can be seen that, in this instance, the ratio of the rate constants falls within a rather narrow range, from 2.5 to 6.1. Similarly, peptides that contain Leu at P2 bind between 100 and 200 times better than the corresponding Ala peptide, whereas Ile peptides bind about 15- to 25-fold worse than the corresponding Leu peptides. In most cases, of the 15 combinations of aa that are compared in Table I, the range in the ratio observed is considerably less than an order of magnitude. The variability in these ratios may be due to experimental error (especially in cases where one or both dissociation rates are <10 min or more than 2000 min) or may be dependent on the other amino acids present in the same peptide (see *Discussion*).

Table I. Comparison of β2m dissociation rates of HLA-A2 complexes containing peptides differing by one amino acid at given position

aa1/aa2[a]	Ratio[b]	Peptide 1		Peptide 2	
		Sequence	$t_{1/2}$[c]	Sequence	$t_{1/2}$[c]
K1/G1	4.6	KALGFVFTL	410	GALGFVFTL	89
K1/G1	2.5	KILGFVFTL	2500	GILGFVFTL	1000
K1/G1	4.8	KILGKVFTL	2100	GILGKVFTL	440
K1/G1	4.3	KLFGGGGGV	470	GLFGGGGGV	110
K1/G1	6.1	KLFGGVGGV	730	GLFGGVGGV	120
L2/A2	>24	GLFGGVGGV	120	GAFGGVGGV	<5
L2/A2	170	GLLGFVFTL	15000	GALGFVFTL	89
L2/I2	>22	GLFGGGGGV	110	GIFGGGGGV	<5
L2/I2	>24	GLFGGVGGV	120	GIFGGVGGV	<5
L2/I2	15	GLLGFVFTL	15000	GILGFVFTL	1000
L2/Q2	11	GLFGGVGGV	120	GQFGGVGGV	11
L2/Q2	37	GLLGFVFTL	15000	GQLGFVFTL	410
L2/Q2	3.4	LLFGYPVYV	6400	LQFGYPVYV	1900
L2/M2	5	GLFGGGGGV	110	GMFGGGGGV	22
L2/M2	1.4	GLFGGVGGV	120	GMFGGVGGV	84
L2/M2	.88	LLFGYPVYV	6400	LMFGYPVYV	7300
A3/E3	10	GIAGFVFTL	240	GIEGFVFTL	24
A3/E3	43	ILASLFAAV	348	ILESLFAAV	8.1
F3/G3	>100	GLFGGGGGV	110	GLGGGGGGV	N/A
F3/G3	>24	GLFGGVGGV	120	GLGGGVGGV	<5
F3/G3	>400	GLFGGGFGV	2000	GLGGGGFGV	<5
F3/G3	>36	GLFGGFGGV	180	GLGGGFGGV	<5
F3/L3	>22	GLFGGGVGV	110	GLLGGGVGV	<5
F3/L3	9.2	GLFGGGVGV	830	GLLGGGVGV	90
G4/K4	4.0	GILGFVFTL	1000	GILKFVFTL	250
G4/K4	1.1	GLFGGGGGV	110	GLFKGGGGV	104
F5/K5	2.3	GILGFVFTL	1000	GILGKVFTL	440
F5/K5	1.2	KILGFVFTL	2500	KILGKVFTL	2100
V6/G6	1.1	GLFGGVGGV	120	GLFGGGGGV	110
V6/G6	3.8	GMFGGVGGV	84	GMFGGGGGV	22
V6/G6	1.6	KLFGGVGGV	730	KLFGGGGGV	470
F7/A7	10.3	GILGFVFTL	1000	GILGFVATL	97
F7/A7	2.6	GLFGGGFGV	2000	GLFGGGAGV	770
F7/E7	15.0	GILGFVFTL	1000	GILGFVETL	65
F7/E7	3.8	GLFGGGFGV	2000	GLFGGGEGV	530
V7/G7	7.5	GLFGGGVGV	830	GLFGGGGGV	110
V7/G7	>18	GLLGGGVGV	90	GLLGGGGGV	<5
L9/A9	1.8	GILGFVFTL	1000	GILGFVFTA	550
L9/A9	>4.8	GLFGGGGGL	24	GLFGGGGGA	<5

[a] aa, using single letter code, and peptide position to be compared.
[b] $t_{1/2}$ of the first peptide divided by $t_{1/2}$ *of the second peptide.*
[c] Half-life of β_2m dissociation for HLA-A2 complexes containing the peptide, in min at 37°C.

Prediction of β_2m dissociation rates using coefficients specific for each combination of aa and peptide position

When a given peptide binds to a class I molecule, the stability of the complex can theoretically be divided into coefficients that represent the contribution of each aa within the peptide to the overall stability. To the first approximation, each of these coefficients could be independent of the sequence of the rest of the peptide; that is, there could be IBS. The ratio in Table I, second column, can be thought of as the ratio of the corresponding coefficients (see Table V). To test the IBS idea, the coefficients must be experimentally determined for a reasonably large set of peptides. For a 9 aa peptide, there would be 180 (9 residues × 20 aa) possible coefficients. For our initial cal-

culations, we chose to limit the number of variables by solving for only those coefficients that are most important for binding to HLA-A2 (82 coefficients in all), based on the peptides that we have studied. All other coefficients were assigned a neutral value of 1.0.

We calculated these coefficients from data on the stability of HLA-A2 complexes containing individual peptides, and also took account of which peptides did not bind to HLA-A2. Peptides that form HLA-A2 complexes can be distinguished from nonbinding peptides by use of a GF assay in which the ability of each peptide to promote incorporation of ^{125}I-β_2m into HLA-A2 complexes is assessed. Average percentages of peptide-dependent β_2m incorporation are listed in column 2 of Tables II, III, IV, and VI. The peptides that formed HLA-A2 complexes could be further subdivided according to how stable the complexes were once they were formed. Those peptides that formed complexes that had a half-life of dissociation of >5 min at 37°C (see column 3) are listed in Table II, whereas those peptides that formed less stable complexes are listed in Table III. Peptides that caused the incorporation of less than 10% of the labeled β_2m into HLA-A2 complex when assayed at a concentration of 1 mM were considered to be *nonbinders* for HLA-A2, and are listed in Table IV. In the calculations, each peptide in Table II corresponded to one independent equation; in which the product of the appropriate coefficients and an overall normalization constant was set equal to the experimentally measured half-life (see *Materials and Methods*). In contrast, the peptides in Table III and Table IV corresponded to an inequality; in which the product of the appropriate coefficients and the overall normalization constant was set equal to less than a half-life of 5 min. For the purpose of discussion, each category of peptide was further divided by sequence into three categories; those based on a poly-Gly or poly-Ala backbone (Tables IIA, IIIA, and IVA), those related to the M1 peptide, which is an optimal HLA-A2 restricted antigenic influenza A matrix peptide (Table IIB and IVB), and other peptides (Table IIC, IIIB, and IVC), including HLA-A2-restricted antigenic peptides and essentially random viral peptides that we had synthesized.

When the equations and inequalities corresponding to each of the peptides listed in Tables II, III, and IV were simultaneously solved, the coefficients listed in Table V were obtained. These coefficients were then used to calculate the theoretical half-life of dissociation of β_2m listed in the column labeled "theo" in Tables II, III, and IV. It can be seen from the ratio listed in the fifth column of Table II that, in every case where accurate experimental half-lives are obtainable, the theoretical binding stabilities differ from the actual binding stabilities by less than a factor of 5.0, and the average ratio is a factor of 1.6. The overall fit of the data is shown graphically in Figure 1*A*, where the theoretical half-life of β_2m dissociation is plotted vs the experimental half-life. Considering that these rate constants vary over at least four orders of magnitude, the fit is impressive. A ratio of coefficients similar to that listed in Table I, second column, can be calculated from the coefficients in Table V. For example, the ratio of coefficients from Table V for K1/G1 is 3.465/0.578 = 6.0, compared to a range of between 2.5 and 6.1 as shown in Table I, second column. This verifies that the coefficients shown in Table V faithfully reflect the contribution of each aa of a nonamer peptide for binding to HLA-A2. The fact that the majority of the half-lives is predicted well (Fig. 1*A*) supports the premise that side-chain/side-chain interactions are in the majority of cases of minimal importance in peptide binding.

Two features of the coefficients listed in Table V are of particular relevance. First, the most important coefficients in Table V are those that are significantly different from 1.0. Second, the higher the frequency of the coefficient among the equations, the more accurately known the value of the coefficient. This is because the value of a coefficient has a greater impact on the overall error if the coefficient is present in a large number of equations, especially if the peptides that correspond to those equations form stable HLA-A2 complexes. The frequency of each aa/peptide position combination in peptides that form stable HLA-A2 complexes (Table II) and in peptides that do not form stable complexes (Tables III and IV) is listed in parentheses in Table V. It can be seen that the coefficient for L2 is both important and accurately known, because it has a high numerical value (103.183), and it appears in 33 different equations (corresponding to 33 peptides that form stable HLA-A2 complexes), and in 39 additional inequalities (corresponding to 39 different peptides that either form unstable HLA-A2 complexes, or do not bind at all). In contrast, the coefficient for K2 is more tentative, because only one peptide containing a Lys at P2 was tested that formed a stable HLA-A2 complex[3]. To get the most accurate values for the coefficients, we included in the calculations equations corresponding to as large a number of peptides as possible, because each additional peptide adds an additional constraint to the values of nine different coefficients. However, there were certain peptides that were excluded from the set (listed in Table VI) because their binding properties appeared to be inconsistent with the bulk of the peptides. In particular, three peptides stood out (Table VIA) that bound reasonably well to HLA-A2 and formed complexes in high yield. These three peptides, we believe, violate the assumption of independent binding of side-chains (see below). The remainder of the peptides were excluded because the data seemed in some way to be "dubious" (Table VIB). Some of these peptides promoted the incorporation of a rather small percentage of ^{125}I-β_2m into complexes. One possible explanation for this could be that a contaminant in the peptide preparation is the active

[3] Analysis of the stability of complexes containing five additional peptides that contain Lys at P2 indicates that the coefficient for Lys at P2 as listed in Table V is too high, and should be close to 1.0.

Table II. A. Poly-Gly and poly-Ala nonapeptides that form stable HLA-A2 complexes

Sequence[a]	GF[b]	Expt. $t_{1/2}$[c]	Theoretical		Predicted	
			$t_{1/2}$[d]	Ratio[e]	$t_{1/2}$[f]	Ratio[g]
ALFAAAAAV	70	570	870	1.5	1100	2.0
GIFGGVGGV	70	8	11	1.4	12	1.6
GLDKGGGGV	70	6	3	2.4	0.6	10
GLFGGFGGV	80	180	190	1.0	180	1.0
GLFGGGAGV	80	770	430	1.8	270	2.9
GLFGGGEGV	90	530	350	1.5	190	2.9
GLFGGGFGV	90	2000	2700	1.3	3300	1.6
GLFGGGGGL	70	24	22	1.1	21	1.2
GLFGGGGGV	70	110	45	2.5	39	2.8
GLFGGGVGV	90	830	480	1.7	230	3.6
GLFGGVGGV	80	120	110	1.1	110	1.2
GLFGGVGKV	80	96	110	1.1	110	1.2
GLFKGVGGV	80	100	110	1.1	110	1.1
GLGGGGFGV	70	8	21	2.8	120	16
GLLGGGVGV	90	90	160	1.7	300	3.4
GLYGGGGGV	60	140	30	4.5	4.1	33
GMFGGGGGV	90	22	25	1.1	27	1.2
GMFGGVGGV	70	84	61	1.4	50	1.7
GQFGGVGGV	70	8	11	1.3	13	1.6
GVFGGVGGV	60	6	6	1.0	1.1	6.4
KLFGGGGGV	90	470	270	1.7	210	2.2
KLFGGVGGV	80	730	660	1.1	630	1.2
B. M1-related nonapeptides that form stable HLA-A2 complexes						
AILGFVFTL	80	1800	1300	1.4	1100	1.6
GAIGFVFTL	80	38	38	1.0	20	1.9
GALGFVFTL	50	89	76	1.2	59	1.5
GELGFVFTL	60	280	280	1.0	180	1.5
GIAGFVFTL	50	240	210	1.1	220	1.1
GIEGFVFTL	60	21	11	1.8	4.3	4.9
GILAFVFTL	40	610	770	1.3	760	1.2
GILGAVFTL	50	220	130	1.7	78	2.8
GILGEVFTL	80	220	96	2.3	14	15
GILFGAFTL	60	220	300	1.4	360	1.6
GILGFEFTL	80	730	960	1.3	2100	2.9
GILGFKFTL	50	71	71	1.0	300	4.2
GILGFVATL	50	97	120	1.2	130	1.3
GILGFVETL	70	65	99	1.5	180	2.7
GILGFVFAL	40	1000	770	1.3	740	1.4
GILGFVFEL	90	870	770	1.1	750	1.2
GILGFVFKL	90	1200	720	1.5	730	1.6
GILGFVFTA	70	560	330	1.7	30	19
GILGFVFTL	50	1000	770	1.3	760	1.4
GILGFVFVL	90	210	230	1.1	310	1.5
GILGFVKTL	50	68	73	1.1	93	1.4
GILGKVFTL	70	440	270	1.7	170	2.6
GILKFVFTL	80	250	770	3.1	820	3.3
GILPFVFTL	90	1100	770	1.5	750	1.5
GIVGFVFTL	80	180	450	2.5	940	5.2
GKLGFVFTL	30	1500	1600	1.0	75	21
GLLGFVFTL	80	15000	7800	2.0	6000	2.6
GQLGFVFTL	90	410	760	1.9	1500	3.6
KALGFVFTL	90	380	450	1.2	540	1.4
KILGFVFTL	50	2700	4600	1.7	5600	2.1
KILGKVFTL	90	2100	1600	1.3	1200	1.8
C. Other nonapeptides that form stable HLA-A2 complexes						
AILLGVFML	20	54	110	2.1	440	8.2
AIYKRWIIL	60	7	7	1.0	340	48
ALFFFDIDL	30	67	67	1.0	41000	610
ATVELLSFL	70	6	5	1.3	3.1	2.0
CLFGYPVYV	90	4600	4600	1.0	7700	1.7

[a] Sequence in single-letter aa code.
[b] Average % of β_2m incorporation as assessed by gel filtration.
[c] Experimentally measured half-life of β_2m dissociation in min at 37°C.
[d] Theoretical half-life of β_2m dissociation, calculated using coefficients in Table V.
[e] Factor by which the theoretical half-life differs from the measured half-life: (column 4 ÷ column 3, or column 3 ÷ column 4, whichever is >1).
[f] Predicted half-life of β_2m dissociation using coefficients that were calculated from all of the equations except the equation corresponding to the peptide that is being predicted.
[g] Factor by which the predicted half-life differs from the measured half-life: (column 6 ÷ column 3, or column 3 ÷ column 6, whichever is >1).

Table II—Continued

Sequence[a]	GF[b]	Expt. $t_{1/2}$[c]	Theoretical $t_{1/2}$[d]	Theoretical Ratio[e]	Predicted $t_{1/2}$[f]	Predicted Ratio[g]
C. Other nonapeptides that form stable HLA-A2 complexes (*Continued*)						
FIFPNYTIV	90	200	300	1.5	670	3.4
IISLWDQSL	60	6	6	1.1	9.8	1.7
ILASLFAAV	70	350	330	1.0	340	1.1
ILESLFAAV	70	10	18	1.8	45	4.7
KLGEFFNQM	90	350	190	1.8	44	7.9
KLGEFYNQM	80	220	150	1.5	57	3.9
KMFGYPVYV	90	4400	15000	3.4	58000	13
LLFGYPVYV	70	6400	7800	1.2	9000	1.4
LLWKGEGAV	80	360	270	1.3	120	3.1
LMFGYPVYV	90	7300	4400	1.7	3100	2.4
LNFGYPVYV	40	41	41	1.0	350	8.6
LQFGYPVYV	90	1800	750	2.5	270	6.9
NIVAHTFKV	80	140	100	1.4	75	1.9
NLVPMVATV	90	480	430	1.1	360	1.3
QMLLAIARL	80	110	49	2.3	13	8.5
QMWQARLTV	90	420	560	1.3	1400	3.3
RLLQTGIHV	90	300	330	1.0	280	1.1
RLVNGSLAL	70	81	64	1.3	36	2.3
SLYNTVATL	80	370	720	1.9	2100	5.6
TLNAWVKVV	70	100	94	1.1	70	1.4
WLYRETCNL	80	91	210	2.3	1600	18
YLFKRMIDL	90	570	420	1.4	390	1.5

species, as has been found in other assay systems (19). In most cases, dissociation rates for the peptides listed in Table VIB were difficult to calculate because the majority of the counts incorporated into HLA-A2 complex dissociated rapidly, although a small percentage of the complex dissociated with the half-life listed. For ALFAAAAAY and GQLGFVFTK no internally consistent half-life could be obtained. In addition, all of these peptides have anchor residues at P2, P3, or P9 (marked in bold in Table VIB), that are infrequent or absent among peptides that form stable HLA-A2 complexes. For all of these reasons, we believe it would be wisest to exclude these peptides from the calculations for the time being, until more peptides are synthesized and tested that could help address these problems.

Explanation of the values obtained for the coefficients

The coefficients in Table V corroborate the data obtained previously from endogenous peptide sequence analyses (5, 8) that the most important anchor positions are at P2 and P9. In addition to Leu and Met at P2, Ile and Gln are also relatively well tolerated. Although Gln has not been previously reported to be an anchor residue at P2 for wild-type HLA-A2, it was recently found to be present at P2 in pooled endogenous peptides isolated from mutant HLA-A*0205 molecules (20), which differ from A*0201 molecules by a single substitution (F9Y) in the B pocket. Our data suggest that the most abundant anchor residues at P2 for this mutant are different from wild type HLA-A2 in a quantitative sense only. At P1, negatively charged residues are unfavorable, whereas Lys is favorable. This can most easily be explained by an ionic interaction with E63, which is known to be located near the N-terminus of the peptide-binding site (4). At P3, aromatic residues are favorable, and charged residues are most often unfavorable. A few exceptional peptides, notably ILDKKVEKV and ILKEPVHGV, can form stable HLA-A2 complexes despite the charged residue at P3, presumably by means of overriding favorable interactions with other peptide residues (a violation of the IBS condition). Most residues are equally well tolerated at P4; however, our data tentatively indicate that large hydrophobic residues like Phe are unfavorable. At P5–P7, aromatic residues seem to be favored, as at P3; however, KLFGFVFTV, which contains Phe at P3, P5, and P7, binds much less well than would be predicted (2,000 min vs 300,000 min predicted) if each of these positions contributed independently. Most likely, this is due to the limited space that is available within the peptide-binding groove to accommodate bulky side-chains (this would be a second violation of the IBS principle). At P8, Val is significantly less favorable than Ala, Glu, Lys, or Thr, at least in the context of the matrix peptide sequence (GILGFVFTL). This may indicate that the hydrophobic isopropyl group of Val cannot be accommodated as easily as hydrophilic, or smaller side-chains. At P9, Val and Leu are better than Met and Ile, and all other residues examined appear to be very much worse. The importance of the P9 position is exemplified by the data collected using peptides that belong to the paradigm GLFGGGFGX, because GLFGGGFGF, GLFGGGFGN, and GLFGGGFGS form complexes that are at least 1000-fold less stable than GLFGGGFGV. Moreover, most peptides that contain either Lys or Tyr at P9 do not bind appreciably, despite

Table III. A. Poly-Gly and poly-Ala nonapeptides that form unstable HLA-A2 complexes

Sequence[a]	GF[b]	$t_{1/2}$[c]	Theo[d]	Ratio[e]
GAFGGVGGV	20		1.1	
GAFGGVGGY	30		0.002	
GEFGGVGGV	20		4.0	
GGFGGVGGV	10		0.10	
GIFGGGGGV	40	3	4.4	1.3
GIGGFGGGL	20		0.12	
GIGGGGGGL	20		0.016	
GLDGGGGGV	60	4	2.9	1.5
GLDGKGGGV	10		7.4	
GLDKKGGGV	30	1	7.4	7.4
GLFGGGFGF	70	5	4.9	1.1
GLFGGGFGG	50	1	4.9	4.9
GLFGGGFGN	80	3	4.9	1.5
GLFGGGFGS	60	1	4.9	4.9
GLFGGGGGA	50	1	9.2	6.6
GLFGGGGGI	70	4	4.9	1.1
GLFGGGGGM	60	4	14	3.1
GLFGGGGGT	50		0.082	
GLFGGGGGY	10		0.083	
GLGFGGGGV	40	2	0.009	250.0
GLGGFGGGV	60	5	2.6	1.8
GLGGGFGGV	60	3	1.4	2.3
GLGGGGGFV	60	3	0.34	7.5
GLGGGGGGY	10		0.001	
GLGGGVGGV	40	1	0.084	1.1
GLLGGGGGV	50	3	15	5.2
GLPGGGGGV	40		5.0	
GNFGGVGGV	10		0.58	
GSFGGVGGV	20		5.0	
GTFGGVGGV	40	1	6.4	4.3
B. Other nonapeptides that form unstable HLA-A2 complexes				
AGNSAYEYV	10		0.21	
GLFPGQFAY	10		4.8	
HILLGVFML	10		5.0	
ILESLFRAV	20	2	5.0	2.7
KKKYKLKHI	10		0.18	
MLASIDLKY	20		0.14	
MLERELVRK	10		0.011	

[a] Sequence, in single-letter aa code.
[b] Average % of β_2m incorporation as assessed by gel filtration.
[c] Experimentally measured half-life of β_2m dissociation in min at 37°C. If no number is present, the half-life was difficult to measure, but is probably less than 5 min.
[d] Theoretical half-life of β_2m dissociation, calculated using coefficients in Table V.
[e] Factor by which the theoretical half-life differs from the measured half-life.

Table IV. A. Poly-Gly and poly-Ala nonapeptides that do not bind to HLA-A2

Sequence[a]	GF[b]	Theo[c]
ALAAAAAAK	1	0.24
GDFGGVGGV	4	5.0
GFFGGVGGV	5	5.0
GHFGGVGGV	8	5.0
GIFGGGGGA	9	0.90
GIGGGFGGL	3	0.068
GIGGGGFGL	2	1.0
GLFGGGGGF	6	0.08
GLGGGGGGL	4	0.17
GLGGGGGGV	5	0.34
GPFGGVGGV	5	5.0
GRLGGGGGI	5	0.036
GYFGGVGGV	6	5.0
B. M1-related nonapeptides that do not bind to HLA-A2		
EILGFVFTK	0	5.0
GILGFVFTE	2	4.9
GILGFVFTK	4	5.0
C. Other nonapeptides that do not bind to HLA-A2		
DIYRIFAEL	4	5.0
EIKDTKEAL	3	0.01
EIYKRWIIL	7	4.0
ELDAPNSHY	1	0.059
ELKSKYWAI	1	5.1
ELKVKNLEL	2	1.1
ELRSLYNTV	2	5.0
ELRSRYWAI	3	3.3
ERYLKDQQL	4	4.9
GEIYKRWII	5	4.0
GLPVGGNEK	5	0.15
GMQWNSTAF	4	0.044
ILKQKIADL	8	1.9
ILRGSVAHK	7	0.020
KIFIAGNSA	1	5.0
KLYEKVYTY	3	5.0
LGFVFTLTV	5	5.0
LLSFLPSDF	5	0.004
PLNPFVSHK	1	3.7
RYWAIRTRS	2	0.082
TPQDLNTML	3	1.7

[a] Sequence, in single letter aa code.
[b] Average % of β_2m incorporation, as assessed by gel filtration.
[c] Theoretical half-life of β_2m dissociation, calculated using coefficients in Table V.

otherwise very favorable residues (e.g., GILGFVFTK, KLYEKVYTY; see Table IVB and IVC).

Application of the binding coefficients to ranking of known antigenic and endogenous peptides

It would be interesting to know if the known endogenously synthesized self and antigenic peptides are among the best HLA-A2 binding peptides. Theoretically, large numbers of peptides may be more capable of binding to HLA-A2, but might never be generated in vivo. To determine whether this is likely, the coefficients in Table V were used to rank all of the potential nonamers from each of the proteins for which a known antigenic or endogenous peptide has been identified. The parameters that describe the ranking of each peptide are shown in Table VII. Column 3 shows the number of overlapping nonamers that could be generated from each protein. Column 4 shows the theoretical half-life of β_2m dissociation for the most stable nonamer. The next two columns list the rank of the peptide using the experimentally measured half-life of dissociation, followed by the measured half-life. Finally, the last two columns list the rank of peptide when the peptide's theoretical half-life of dissociation is used, followed by the theoretical half-life. (Note that our current algorithm is capable of ranking nonamers only, although some longer peptides could form comparably stable complexes.) The first peptide, the influenza matrix peptide GILGFVFTL, was previously found to be a major target of all HLA-A2-restricted, influenza-specific CTL, both in humans (21) and in HLA-A2 transgenic mice (22). It ranks first among all possible nonamers from the matrix protein

Table V. Coefficients used to calculate theoretical rate constants

aa[a]	Coeff[b]	Freq[c]	aa	Coeff	Freq	aa	Coeff	Freq
A1	1.000	(6, 2)	A4	1.000	(4, 3)	A7	1.000	(8,4)
C1	0.597	(1, 0)	C4	1.000	(0, 0)	C7	1.000	(1, 0)
D1	0.041	(0, 1)	D4	1.000	(0, 2)	D7	1.000	(0, 0)
E1	0.578	(0, 9)	E4	1.000	(3, 1)	E7	0.820	(2, 2)
F1	1.000	(1, 0)	F4	0.027	(1, 2)	F7	6.383	(32, 10)
G1	0.578	(46, 48)	G4	1.000	(53, 44)	G7	0.105	(16, 37)
H1	0.044	(0, 1)	H4	1.000	(0, 0)	H7	1.000	(0, 0)
I1	1.000	(3, 3)	I4	0.078	(0, 1)	I7	1.000	(4, 1)
K1	3.465	(8, 3)	K4	1.000	(6, 2)	K7	0.603	(2, 1)
L1	1.000	(5, 2)	L4	0.646	(3, 2)	L7	1.000	(2, 3)
M1	1.000	(0, 2)	M4	1.000	(0, 0)	M7	1.000	(0, 0)
N1	1.000	(2, 0)	N4	1.000	(2, 0)	N7	1.000	(2, 3)
P1	1.000	(0, 1)	P4	1.000	(3, 2)	P7	1.000	(0, 0)
Q1	1.000	(2, 0)	Q4	1.000	(2, 1)	Q7	1.000	(1, 1)
R1	1.000	(2, 1)	R4	1.000	(1, 2)	R7	0.277	(0, 1)
S1	1.000	(1, 0)	S4	1.000	(2, 6)	S7	1.000	(1, 3)
T1	1.000	(1, 1)	T4	1.000	(0, 0)	T7	1.000	(1, 3)
V1	1.000	(0, 0)	V4	1.000	(0, 3)	V7	1.120	(8, 1)
W1	1.000	(1, 0)	W4	1.000	(0, 1)	W7	5.951	(0, 3)
Y1	1.000	(1, 0)	Y4	1.000	(0, 2)	Y7	1.000	(0, 1)
A2	1.000	(3, 2)	A5	1.000	(4, 3)	A8	1.000	(6, 7)
C2	0.500	(0, 0)	C5	1.000	(0, 0)	C8	1.000	(0, 0)
D2	0.500	(0, 1)	D5	1.000	(0, 0)	D8	1.000	(2, 2)
E2	2.840	(1, 2)	E5	0.756	(2, 1)	E8	1.000	(1, 3)
F2	0.500	(0, 1)	F5	6.044	(30, 7)	F8	1.000	(1, 1)
G2	0.500	(0, 3)	G5	0.804	(24, 41)	G8	1.000	(20, 41)
H2	0.500	(0, 1)	H5	1.000	(1, 0)	H8	1.000	(1, 4)
I2	10.151	(30, 14)	I5	1.000	(0, 3)	I8	1.000	(2, 2)
K2	20.524[3]	(1, 1)	K5	2.085	(2, 9)	K8	1.000	(3, 1)
L2	103.183	(33, 39)	L5	1.000	(3, 4)	L8	1.000	(0, 0)
M2	57.920	(6, 1)	M5	1.000	(1, 0)	M8	1.000	(1, 2)
N2	0.542	(1, 1)	N5	1.000	(1, 1)	N8	1.000	(1, 0)
P2	0.500	(0, 2)	P5	1.000	(0, 1)	P8	1.000	(0, 0)
Q2	10.006	(3, 0)	Q5	1.000	(0, 0)	Q8	1.000	(2, 1)
R2	0.500	(0, 2)	R5	1.000	(2, 2)	R8	1.000	(1, 2)
S2	0.500	(0, 1)	S5	1.000	(0, 1)	S8	1.000	(1, 1)
T2	6.080	(1, 1)	T5	1.000	(2, 1)	T8	1.000	(30, 6)
V2	5.919	(1, 0)	V5	1.000	(0, 0)	V8	0.293	(2, 0)
W2	0.500	(0, 0)	W5	2.680	(2, 0)	W8	1.000	(0, 0)
Y2	0.500	(0, 2)	Y5	8.002	(6, 0)	Y8	1.000	(6, 1)
A3	1.000	(2, 2)	A6	1.000	(2, 1)	A9	1.000	(1, 3)
C3	1.000	(0, 0)	C6	1.000	(0, 0)	C9	0.010	(0, 0)
D3	0.726	(1, 4)	D6	1.000	(2, 2)	D9	0.010	(0, 0)
E3	0.054	(2, 2)	E6	3.246	(2, 0)	E9	0.015	(0, 1)
F3	11.383	(27, 27)	F6	4.369	(4, 4)	F9	0.009	(0, 4)
G3	0.088	(3, 12)	G6	1.060	(13, 29)	G9	0.009	(0, 1)
H3	1.000	(0, 0)	H6	1.000	(0, 0)	H9	0.010	(0, 0)
I3	1.849	(1, 1)	I6	1.000	(1, 1)	I9	0.534	(0, 6)
K3	0.024	(0, 5)	K6	0.239	(1, 1)	K9	0.015	(0, 7)
L3	3.685	(31, 6)	L6	1.000	(1, 2)	L9	2.357	(41, 13)
M3	1.000	(0, 0)	M6	1.000	(1, 0)	M9	1.501	(2, 1)
N3	1.000	(1, 2)	N6	1.000	(0, 3)	N9	0.009	(0, 1)
P3	1.261	(0, 2)	P6	1.000	(6, 1)	P9	0.010	(0, 0)
Q3	1.000	(0, 2)	Q6	1.000	(0, 1)	Q9	0.010	(0, 0)
R3	0.033	(0, 3)	R6	1.000	(1, 2)	R9	0.010	(0, 0)
S3	1.000	(1, 1)	S6	1.000	(1, 1)	S9	0.009	(0, 2)
T3	1.000	(0, 0)	T6	1.000	(2, 1)	T9	0.009	(0, 1)
V3	2.173	(5, 0)	V6	2.588	(40, 20)	V9	4.884	(36, 27)
W3	12.978	(2, 1)	W6	0.251	(1, 1)	W9	0.010	(0, 0)
Y3	7.613	(4, 4)	Y6	3.470	(2, 4)	Y9	0.009	(0, 7)

[a] aa using single letter code, followed by the position within the peptide.

[b] Coefficient, calculated by solving simultaneously equations corresponding to each of the peptides in Tables II, III, and IV. Coefficients whose value equal exactly 1.000 were constrained to equal 1.0. No coefficient is known to better than two decimal places; many coefficients may be off by greater than a factor of 2.0. The value in this table is representative of the raw output from the Fortran program. At P2, coefficients were assigned a value of 0.500 if no peptides were studied that formed stable complexes with this aa/peptide position combination. At P9, coefficients were assigned a value of 0.010 if no peptides were studied that contained this aa/peptide position combination. Note that all undetermined coefficients in Table V have been assigned the value of 1.0, which corresponds to the coefficient for Ala at that same position. In making predictions of the stability of HLA-A2 complexes containing unknown peptides, one could substitute a coefficient with the corresponding coefficient of a chemically more similar aa. The overall normalization coefficient = 0.151.

[c] First number: number of peptides that contain the aa at the position in question in Table II. Second number: number of peptides that contain the aa at the position in question in Tables III and IV.

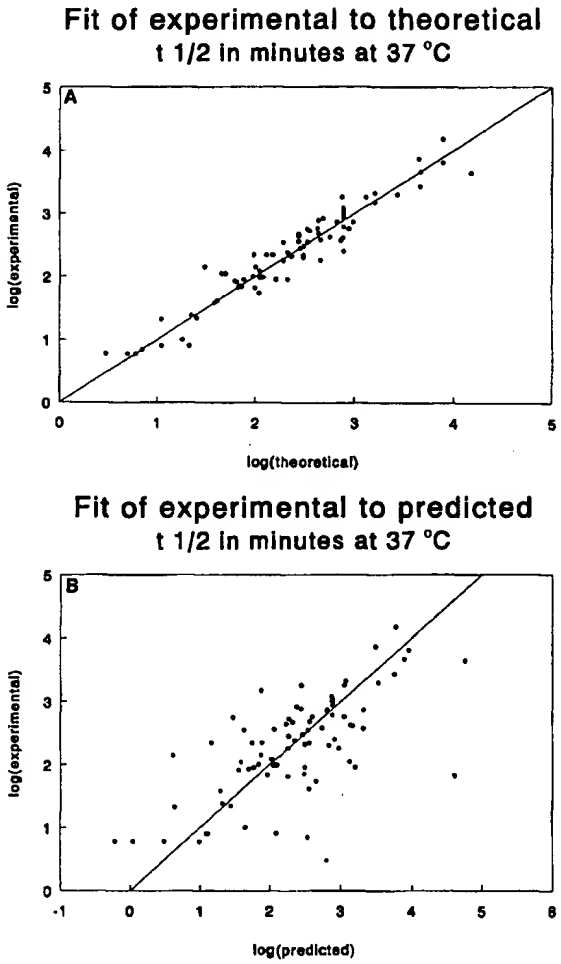


FIGURE 1. *A*, Comparison of theoretical half-life of β_2m dissociation to the experimentally measured half-life. The data from the third column of Table II were plotted against the fourth column. The *line* indicates the position of a perfect fit. *B*, Comparison of the predicted half-life of β_2m dissociation to the experimentally measured half-life. The data from the third column of Table II were plotted against the sixth column. The *line* indicates the position of a perfect fit.

(Table VIIA) and third among all possible nonamers encoded by the influenza genome (data not shown). The HTLV-1 derived, HLA-A2-restricted peptide LLFGYPVYV (23) also ranks first from its source protein. The HIV polymerase-derived, HLA-A2-restricted, antigenic peptide ILKEPVHGV (11) "theoretically" ranks 45th of the 1007 possible nonamers in the pol protein, which would place it only in the top 5%. However, in the case of ILKEPVHGV, the experimental rank is much higher than the theoretical rank, because ILKEPVHGV binds much better than expected based on the coefficients in Table V. Notably, none of the other seven higher-ranking HIV polymerase-encoded peptides are predicted to bind much more than twofold better (data not shown). The remaining three antigenic peptides, KLGEFYNQMM (24), FIAGNSAYEYV (23), and FLPSDFFPSV (25), are longer than nine aa long, which is why no theoretical rank of half-life is listed. When the experimentally measured half-life of these peptides is compared against the theoretical half-lives of all possible nonamers from the source protein, each of these peptides ranks close to the top.

When the endogenous peptides are examined, we see that ILDKKVEKV, like ILKEPVHGV, binds much more tightly than expected using the coefficients in Table V. When its experimentally measured half-life is used for ranking purposes, it ranks at the top of the list. With the exception of LLDVPTAAV, the other endogenous peptides also rank in the top few percent of all possible nonamers from their source protein. Note that our estimates for the ranking of these remaining endogenous peptides are inherently less accurate because we have not measured the half-lives of complexes containing these peptides. The endogenous peptide that ranks the lowest, LLDVPTAAV, was derived from the leader peptide of IP30, and was isolated from a cell line with a mutation in Ag processing, so that presumably only peptides derived from the leader peptide were available for binding to HLA-A2 (26). We conclude that most antigenic peptides and most predominant self peptides are selected from among those peptides that can form the most stable class I complexes.

Table VI. A. Nonapeptides that may violate the side-chain independence rule

Sequence[a]	GF[b]	$t_{1/2}$[c]	Theo[d]	Ratio[e]
KLFGFVFTV	60	2000	300,000	150
ILDKKVEKV	50	2900	250	12
ILKEPVHGV	80	190	4.8	38
B. Peptides that form HLA-A2 complexes that behave irregularly				
ALFAAAAAY	30		2	
GIGFGGGGL	20	200	0	200000
GKFGGVGGV	10	80	22	3.7
GLFGGGGGK	30		0	
EILGFVFTL[f]	10	85	770	9.1
GIKGFVFTL[g]	60	2500	5	500
GQLGFVFTK	70		5	
ILGFVFTLT[h]	50	140	0	
KILGFVFTK	5	210	30	7.2
KKLGFVFTL	30	750	9300	13
KLFEKVYNY	20	9	8	1.2
LRFGYPVYV	20	400	180	2.3

[a] Sequence, in single letter aa code.
[b] Average % of β_2m incorporation, as assessed by gel filtration.
[c] Experimentally measured half-life of β_2m dissociation in min at 37°C.
[d] Theoretical half-life of β_2m dissociation, calculated using coefficients in Table V.
[e] Factor by which the theoretical half-life differs from the measured half-life.
[f] This peptide would have been placed in Table IIB if it were better able to form HLA-A2 complexes. The dissociation rate of complexes containing this peptide is consistent with the rest of the data, but it was not used to calculate the coefficients.
[g] This peptide reproducibly fails to form complexes with the expected pI. Instead, the complex has the same charge as GILGFVFTL complexes.
[h] This peptide is likely to be contaminated with trace amounts of ILGFVFTL, which is known to form complexes with the measured stability (16).

Table VII. Ranking of HLA-A2 antigenic and endogenous peptides using the coefficients[a]

Protein[b]	Sequence[c]	No. 9-mers[d]	Highest $t_{1/2}$[e]	Experimental[f]		Theoretical[g]	
				Rank	$t_{1/2}$	Rank	$t_{1/2}$
A. Antigenic peptides							
Flu matrix (10, 21)	GILGFVFTL	244	800	1st	1000	1st	800
HTLV-1 tax (23)	LLFGYPVYV	350	8000	1st	4000	1st	8000
HIV polymerase (11)	ILKEPVHGV	1007	400	8th	190	45th	10
Influenza nucleoprotein (24)	KLGEFYNQMM	552	600	4th	190		
HCMV gB (23)	FIAGNSAYEYV	889	2000	3rd	1000		
Hepatitis core Ag (25)	FLPSDFFPSV	175	400	1st	1500		
B. Endogenous peptides							
hsp 84 (16)	ILDKKVEKV	715	900	1st	2800	19th	20
ip30 (8)	LLDVPTAAV	295	500			7th	60
tis 21 (8)	TLWVDPYEV	150	1000			1st	1000
helicase (8)	YLLPAIVHI	546	400			8th	30
pp61 (8)	SLLPAIVEL	581	900			6th	200
phosphorylase regulatory A (8)	SLLPAIVEL	581	900			6th	200
phosphorylase regulatory B (8)	SLLPAIVEL	567	900			6th	200

[a] There are other examples of sequences that are known to contain HLA-A2-restricted peptides but, in these other cases, the optimal peptide has not been identified (12–14).
[b] Protein of origin.
[c] Amino acid sequence.
[d] Number of nonamers in the protein.
[e] Theoretical half-life of β_2m dissociation (in min at 37°C) for the peptide that ranked first for this protein.
[f] Rank of peptide determined by comparing the experimentally measured half-life of β_2m dissociation to the theoretical half-life of β_2m dissociation for all the nonamers that could be generated from the same protein. These columns are blank for peptides that have not been tested.
[g] Rank of peptide determined by comparing the theoretical half-life of β_2m dissociation to the theoretical half-life of β_2m dissociation for all the nonamers that could be generated from the same protein. These columns are blank for peptides that are longer than nonamers, because we cannot make an accurate prediction of the theoretical half-life of β_2m dissociation for longer peptides.

Discussion

One of the major reasons to study peptide binding to class I molecules is to be able to determine which peptides are likely to be antigenic, starting from the primary sequence of (for example) a viral protein. In addition, it would be useful to know why certain peptides are antigenic, but most peptides are immunologically silent. The data in Table VII suggests that so far as we can tell, dominant antigenic peptides in HLA-A2-restricted immune responses are among those peptides that bind most tightly to HLA-A2. If this turns out to be generally correct, then it should be possible to develop mathematical algorithms to identify most antigenic peptides using approaches similar to that described herein that are tailored to the peptide-binding properties of each histocompatibility Ag.

The class I MHC protein HLA-A2 has been shown to bind certain peptides, generally 9 aa in length, that preferentially contain Leu or Met at P2 and a Val or Leu at the C-terminus (P9) (5, 8). The residues at these two positions have been termed anchor residues (5) because their relative lack of variability indicates that they serve as primary contact points between the peptide and the class I binding site. However, peptides that contain both Leu at P2 and Val at P9 form complexes whose stability spans at least four orders of magnitude (16), indicating that the aa at other positions can serve as auxiliary anchor residues that are critical for peptide binding. Therefore, to make useful predictions about peptide binding affinity, if possible, the contribution of both the dominant and auxiliary anchor residues must be analyzed on a quantitative basis. The simplest approach is to assume that each amino acid side-chain binds independently of the rest of the peptide (the IBS hypothesis). It seems reasonable to expect that for many peptides, the IBS hypothesis will adequately explain peptide binding, and for other peptides, more complicated explanations will be needed to explain peptide binding. Whenever IBS is true, the binding affinity of any nonamer can be broken down into nine different coefficients, each of which is dependent only on the identity of the aa and the position within the peptide. Therefore, a table containing 180 different coefficients would contain the information necessary to calculate a probable binding affinity for any possible nonamer.

To calculate the coefficients, we measured the stability of a large number of HLA-A2 complexes containing distinct peptides, as assessed by measuring the rate of β_2m dissociation. We also compiled a list of peptides that were unable to make stable complexes with HLA-A2. To solve for the coefficients, the β_2m dissociation data for each peptide was treated as an independent equation, in which the measured half-life of β_2m dissociation was set equal to the product of the nine coefficients (see *Materials and Methods*). In theory, a sufficiently large set of peptide binding data could be used to solve for all of the coefficients simultaneously. In practice, we calculated values for the coefficients that were most important to our current peptide database. Until every aa at every position has been

tested, we cannot exclude the possibility that other coefficients may also contribute significantly to peptide binding to HLA-A2. Despite these approximations, we found that for the vast majority of the peptides that we have tested, the binding data were consistent with the IBS hypothesis. Only for 3 of 83 peptides was it necessary to propose significant side-chain/side-chain interactions to explain the observed peptide-binding properties. We conclude that for most peptides, the stability of the HLA-A2/β_2m/peptide complex is what would be expected if each side-chain of the peptide bound independently to the class I molecule.

Compiling a table of peptide binding coefficients based on individual peptide side-chains has several powerful advantages. First, the coefficients in Table V incorporate all of the peptide-binding data that we have collected so far, with a few exceptions (see Table VI). The table of coefficients can then be used to estimate the binding stability of HLA-A2 complexes containing an untested peptide. As soon as additional binding data become available, the new data can be used to refine the accuracy of the table of coefficients. Second, the table can be used to make a quantitative prediction about which aa in a given peptide are of primary importance for binding to HLA-A2. For example, in the case of the influenza matrix peptide GILGFVFTL, the Phe residues at P5 and P7 are predicted to be almost as important as the Ile at P2. This information could be used to predict which substitutions in an antigenic peptide might allow it to bind more tightly to HLA-A2. In some cases, a peptide that binds very weakly to HLA-A2 might be converted into a useful vaccine candidate by this means. Third, experiments can be designed to test every coefficient in the table by measuring the stability of HLA-A2 complexes containing peptides that differ at the aa in question. Fourth, whenever the binding of a peptide is badly predicted by the table of coefficients, one would predict that significant side-chain/side-chain interactions are taking place or that some side-chain is oriented in a significantly different direction than usual.

The most obvious way to test the validity of the coefficients in Table V would be to predict the half-lives of β_2m dissociation for complexes formed with a new set of peptides, and then to compare the predictions against experimental measurements. We have not explicitly done this, because we have used all new information to improve the values of the coefficients. Instead, to test the power of this methodology to predict which peptides would make the most stable HLA-A2 complexes, the coefficients were recalculated for each of the 80 peptides that bind stably to HLA-A2, using all of the equations used to calculate the coefficients in Table V except for the equation corresponding to the peptide to be tested. The factor by which the "predicted" half-life of β_2m dissociation differs from the measured half-life is listed in Table II, seventh column. It can be seen that although this factor is always greater than the factor obtained when the peptide to be tested is included in the set of equations (Table II, fifth column), the half-lives of 62 of the 80 peptides were still predicted within a factor of five. In most cases, the poorest predictions can be easily explained. For example, ALFFFDIDL (Table IIC) was predicted poorly because it was the only peptide that formed stable HLA-A2 complexes that contained a Phe at P4. When the equation for ALFFFDIDL was deleted, the program calculated the highest value for the coefficient for Phe at P4 that was consistent with the observation that GLGFGGGGV (Table IIIA) and LLSFLPSDF (Table IVC) do not form stable HLA-A2 complexes. It turns out that this causes the value of the coefficient for F4 to increase from 0.027 (Table V) to 16.7, which is an artificially high value. This happens because GLGFGGGGV and LLSFLPSDF have such poor anchor residues at P3 and P9, respectively that the coefficient for F4 could be as high as 16.7, and these peptides would still not be expected to form stable HLA-A2 complexes.

As a further check on the logic behind the calculations, the coefficients were recalculated allowing all of the 180 coefficients to be variables. This allowed the overall error function to decrease from a value of 22.8 to a value of 12.3. The new set of coefficients was very similar to that in Table V (data not shown), especially for the coefficients that apply to a large number of peptides (like L2 and V9). As would be expected considering the number of variables, certain coefficients were poorly defined. For example, the coefficients for both W1 and C7 were present only in the equation corresponding to the data for WLYRETCNL (see Table II), and other coefficients could not be calculated at all because peptides containing the corresponding aa were not available. Nonetheless, with a sufficiently large set of peptides, these difficulties would be overcome, and all of the coefficients could be simultaneously calculated. Thus, it will not always be necessary to make intuitive choices as to which coefficients should be allowed to deviate from a value of 1.0. However, it would be possible to reduce the number of variable coefficients used in our calculations by two distinct means. First, some of the coefficients that were allowed to be variables were calculated to have values near 1.0 (e.g., D3, P3, L4, E5, G5, G6, E7, and V7), and therefore (in retrospect) need not have been variables. Second, in some cases, chemically similar aa were found to have similar coefficients, even though the algorithm used to calculate the coefficients did not take this into account (e.g., F3, Y3, W3 and F5, Y5, W5), and therefore the number of variables could be reduced by constraining several coefficients to have the same value.

The IBS hypothesis is based on the following theoretical considerations. The logarithm of each coefficient can be thought of as being related to a partial free energy of activation for the process of dissociation of the complex. The partial free energies of activation should be additive, assuming the peptide side-chains bind independently to the HLA H chain, and assuming the rate-limiting step for

the dissociation of each complex is the same. The sign of the logarithm of a coefficient can be either negative or positive, depending on whether the aa makes a favorable or unfavorable contribution to the stability of the complex. The equation $E = -RT \ln K$ converts from energy, which is additive, to the coefficients themselves, which are factors that contribute to the value of the rate constant. Instead of measuring free energies, we have measured exclusively the complex stability, as deduced from the half-life of β_2m dissociation. These kinetic measurements may have some advantages over free energy measurements, because there is no contribution from variations in free peptide solvation, and the uncertain status of the HLA H chain/β_2m dimer. However, whenever two peptide side-chains interact with one another, compete for binding to the same pocket, transmit structural changes to other pockets, or alter the structure of the rate-limiting step for dissociation, the IBS condition would be violated.

The IBS hypothesis is based on the approximation that any one aa at a given peptide position would be able to adopt nearly the same limited set of conformations, regardless of the rest of the peptide's sequence. The currently available crystal structure data indicate that the overall structure of the peptide-binding groove is similar for all peptide complexes (4, 27–32), and adjusts only slightly to different peptides (28). In particular, the conformation of the peptide is constrained by the canonical hydrogen bonds between the class I molecule and the peptide termini at both end of the peptide-binding groove (28, 30, 32). In addition, one might expect that for HLA-A2 the anchor residue at P2 would always be buried in the B pocket (4). These constraints would be expected to limit the potential flexibility of the peptide. The energetics of the conformations of each aa, and its interactions with the peptide-binding groove would determine the values of the coefficients, which might also incorporate side-chain solvation effects in the case of exposed aa, and also secondary effects transmitted to other residues by limitations to the conformational flexibility of the peptide backbone. Theoretically, energy minimization calculations based on crystallographically determined coordinates of one peptide/H chain complex should be able to quantitate the energetic consequences of substitutions in the peptide, making the table of coefficients obsolete. At this point in time, however, these calculations are cumbersome and unreliable, and the table of coefficients can provide a first approximation to the binding properties of an unknown peptide.

Because it appears that longer peptides can loop out in the middle in order to maintain favorable contacts at both termini (28, 31), it would be possible to extend the IBS idea to account for the binding properties of peptides longer than 9 aa. In this case, the coefficients for P1-P4 might be the same as with nonamers, but the coefficients for P6-P9 would apply to P(Ω-4) - PΩ, where Ω stands for the last amino acid in the peptide. We have found that the binding properties of some peptides can be explained adequately in this way (data not shown), but many peptides are predicted very poorly, especially when there is a Gly residue at P2 or P3. So far, the longest peptide that we have tested that appears to bind by the looping-out mechanism is the 15-mer GLFGGGGGVKGGFGV, which contains favorable dominant anchor residues at P2 and PΩ, and also favorable auxiliary anchor residues at P3 and P(Ω-2). Before this extension of the IBS hypothesis will be generally useful, it will be necessary to work out an additional set of rules that takes into account the variety of peptide backbone conformations that can be used to accommodate the looped-out residues.

We have used the coefficients listed in Table V to ask whether the well-studied antigenic and endogenous peptides represent the highest affinity peptides that could be generated from their parent proteins. The calculations listed in Table VII indicate that, so far as we can tell, these biologically important peptides are usually among the top 2% of all possible HLA-A2 binding peptides. For example, the optimal HLA-A2-restricted peptide GILGFVFTL is predicted to bind more tightly to HLA-A2 than any other peptide that can be derived from the influenza matrix protein, even though it contains a relatively unfavorable Ile at P2. Thus, there is no reason to believe that antigenic peptides are preferentially selected from a lower affinity set of peptides, as has been proposed (33). It would be interesting to determine whether any of the other peptides that are predicted to form stable complexes with HLA-A2 are ever antigenic or associated with HLA-A2 in vivo. In this way, one could address the relative importance to antigenicity of peptide binding to HLA-A2 compared to other factors like protein proteolysis, protein turnover, peptide stability, peptide transport, the rate of formation of the complex, and holes in the T cell repertoire. Unlike the dissociation of the HLA-A2 complex, which is a unimolecular process, the processes that affect the rate of formation of the HLA-A2 complex are potentially subject to control mechanisms that may differ between cell types, making it much more difficult to study them. In any case, we believe that the coefficients in Table V provide the best means available so far to identify HLA-A2 binding peptides, whether or not they turn out to be antigenic, immunologically silent, or never formed in vivo. It would be interesting to determine whether a table of coefficients calculated by similar means would be able to improve the predictive power of the motifs that have been elucidated for class II binding peptides (34). Other macromolecular interactions such as ligand/antibody and oligonucleotide/DNA binding protein might also be addressed using a mathematical approach similar to that described here.

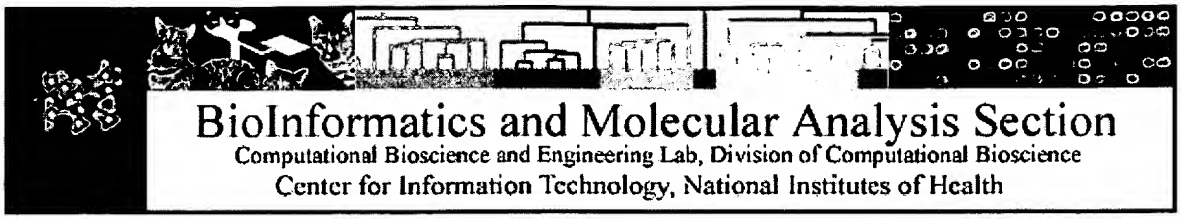
Note added in proof. Software is being developed to make the coefficients in Table V publicly accessible through the National Center for Biotechnology Information at the National Library of Medicine.

Acknowledgments

The authors thank Drs. David Alling, William Biddison, George Hutchinson, William Magner, David Margulies, Michael Shields, and John Spouge for useful discussions, and thank Lisa Hull for excellent technical assistance.

References

1. Bjorkman, P. J., and P. Parham. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem. 59:253.*
2. Townsend, A., and H. Bodmer. 1989. Antigen recognition by class I-restricted T lymphocytes. *Annu. Rev. Immunol. 7:601.*
3. Garrett, T. P., M. A. Saper, P. J. Bjorkman, J. L. Strominger, and D. C. Wiley. 1989. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature 342:692.*
4. Saper, M. A., P. J. Bjorkman, and D. C. Wiley. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol. 219:277.*
5. Falk, K., O. Rötzschke, S. Stevanovic, G. Jung, and H.-G. Rammensee. 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature 351:290.*
6. Jardetzky, T. S., W. S. Lane, R. A. Robinson, D. R. Madden, and D. C. Wiley. 1991. Identification of self peptides bound to purified HLA-B27. *Nature 353:326.*
7. Romero, P., G. Corradin, I. F. Luescher, and J. L. Maryanski. 1991. H-2K^d-restricted antigenic peptides share a simple binding motif. *J. Exp. Med. 174:603.*
8. Hunt, D. F., R. A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A. L. Cox, E. Appella, and V. H. Engelhard. 1992. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectroscopy. *Science 255:1261.*
9. Corr, M., L. F. Boyd, S. R. Frankel, S. Kozlowski, E. A. Padlan, and D. H. Margulies. 1992. Endogenous peptides of a soluble major histocompatibility complex class I molecule, H-2L^d: Sequence motif, quantitative binding, and molecular modeling of the complex. *J. Exp. Med. 176:1681.*
10. Bednarek, M. A., S. Y. Sauma, M. C. Gammon, G. Porter, S. Tamhankar, A. R. Williamson, and H. J. Zweerink. 1991. The minimum peptide epitope from the influenza virus matrix protein: Extra and intracellular loading of HLA-A2. *J. Immunol. 147:4047.*
11. Tsomides, T. J., B. D. Walker, and H. N. Eisen. 1991. An optimal viral peptide recognized by CD8$^+$ T cells binds very tightly to the restricting class I major histocompatibility complex protein on intact cells but not to the purified class I protein. *Proc. Natl. Acad. Sci. USA 88:11276.*
12. Claverie, J., P. Kourilsky, P. Langlade-Demoyen, A. Chalufour-Prochnika, G. Dadaglio, F. Tekaia, F. Plata, and L. Bougueleret. 1988. T-immunogenic peptides are constituted of rare sequence patterns. Use in the identification of T epitopes in the human immunodeficiency virus gag protein. *Eur. J. Immunol. 18:1547.*
13. Clerici, M., D. R. Lucey, R. A. Zajac, R. N. Boswell, H. M. Gebel, H. Takahashi, J. A. Berzofsky, and G. M. Shearer. 1991. Detection of cytotoxic T lymphocytes specific for synthetic peptides of gp160 in HIV-seropositive individuals. *J. Immunol. 146:2214.*
14. Dadaglio, G., A. Leroux, P. Langlade-Demoyen, E. M. Bahraoui, F. Traincard, R. Fisher, and F. Plata. 1991. Epitope recognition of conserved HIV envelope sequences by human cytotoxic T lymphocytes. *J. Immunol. 147:2302.*
15. Choppin, J., J. -G. Guillet, and J. -P. Lévy. 1992. HLA class I binding regions of HIV-1 proteins. *Crit. Rev. Immunol. 12:1.*
16. Parker, K. C., M. A. Bednarek, L. K. Hull, U. Utz, B. Cunningham, H. J. Zweerink, W. E. Biddison, and J. E. Coligan. 1992. Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. *J. Immunol. 149:3580.*
17. Parker, K. C., M. Dibrino, L. Hull, and J. E. Coligan. 1992. The β_2-microglobulin dissociation rate is an accurate measure of the stability of MHC class I heterotrimers and depends on which peptide is bound. *J. Immunol. 149:1896.*
18. Davidon, W. C. 1975. Optimally conditioned optimization algorithms without line searches. *Math. Program 9:1.*
19. Schumacher, T. N. M., M. L. H. De Bruijn, L. N. Vernie, W. M. Kast, C. J. M. Melief, J. J. Neefjes, and H. L. Ploegh. 1991. Peptide selection by MHC class I molecules. *Nature 350:703.*
20. Rötzschke, O., K. Falk, S. Stevanovic, G. Jung, and H.-G. Rammensee. 1992. Peptide motifs of closely related HLA class I molecules encompass substantial differences. *Eur. J. Immunol. 22:2453.*
21. Gotch, F., J. Rothbard, K. Howland, A. Townsend, and A. McMichael. 1987. Cytotoxic T lymphocytes recognize a fragment of influenza virus matrix protein in association with HLA-A2. *Nature 326:881.*
22. Vitiello, A., D. Marchesini, J. Furze, L. A. Sherman, and R. W. Chesnut. 1991. Analysis of the HLA-restricted influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex. *J. Exp. Med. 173:1007.*
23. Utz, U., S. Koenig, J. E. Coligan, and W. E. Biddison. 1992. Presentation of three different viral peptides, HTLV-1 tax, HCMV gB, and influenza virus M1, is determined by common structural features of the HLA-A2.1 molecule. *J. Immunol. 149:214.*
24. Robbins, P. A., L. A. Lettice, P. Rota, J. Santos-Aguado, J. Rothbard, A. J. McMichael, and J. L. Strominger. 1989. Comparison between two peptide epitopes presented to cytotoxic T lymphocytes by HLA-A2. Evidence for discrete locations within HLA-A2. *J. Immunol. 143:4098.*
25. Bertoletti, A., F. V. Chisari, A. Penna, S. Guilhot, L. Galati, G. Missale, P. Fowler, H. -J. Schlicht, A. Vitiello, R. C. Chesnut, F. Fiaccadori, and C. Ferrari. 1993. Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein. *J. Virol. 67:2376.*
26. Henderson, R. A., H. Michel, K. Sakaguchi, J. Shabanowitz, E. Appella, D. F. Hunt, and V. H. Engelhard. 1992. HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science 255:1264.*
27. Madden, D. R., J. C. Gorga, J. L. Strominger, and D. C. Wiley. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell 70:1035.*
28. Fremont, D. H., M. Matsumura, E. A. Stura, P. A. Peterson, and I. A. Wilson. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science 257:919.*
29. Zhang, W., A. C. M. Young, M. Imarai, S. G. Nathenson, and J. C. Sacchettini. 1992. Crystal structure of the major histocompatibility complex class I H-2K^b molecule containing a single viral peptide: Implications for peptide binding and T-cell receptor recognition. *Proc. Natl. Acad. Sci. USA 89:8403.*
30. Silver, M. L., H.-C. Guo, J. L. Strominger, and D. C. Wiley. 1992. Atomic structure of a human MHC molecule presenting an influenza virus peptide. *Nature 360:367.*
31. Guo, H.-C., T. S. Jardetzky, T. P. J. Garrett, W. S. Lane, J. L. Strominger, and D. C. Wiley. 1992. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature 360:364.*
32. Madden, D. R., J. C. Gorga, J. L. Strominger, and D. C. Wiley. 1991. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature 353:321.*
33. Ohno, S. 1992. "Self" to cytotoxic T cells has to be 1000 or less high affinity nonapeptides per MHC antigen. *Immunogenetics 36:22.*
34. O'Sullivan, D., T. Arrhenius, J. Sidney, M.-F. Del Guercio, M. Albertson, M. Wall, C. Oseroff, S. Southwood, S. M. Colón, F. C. A. Gaeta, and A. Sette. 1991. On the interaction of promiscuous antigenic peptides with different DR alleles: identification of common structural motifs. *J. Immunol. 147:2663.*



HLA Peptide Binding Predictions

Function: Rank potential 8-mer, 9-mer, or 10-mer peptides based on a predicted half-time of dissociation to HLA class I molecules. The analysis is based on coefficient tables deduced from the published literature by Dr. Kenneth Parker, Children's Hospital Boston (email: kenneth.parker@childrens.harvard.edu).
Another web site for predicting which peptides bind to MHC molecules is SYFPEITHI, developed by Hans-Georg Rammensee's lab.

Analysis Options: **HLA molecule** **n-mers**

A1
A_0201
A_0205
A24
A3

9

Results Limited by: ◉ **Explicit Number** ○ **Predicted $T_{(½)}$ >=**

20 100

Please enter or paste an AA sequence to analyze (most formats accepted):

☑ Echo input sequence (generally recommended)

submit reset

Credits: WWW implementation by Ronald Taylor of BIMAS / CBEL / CIT / NIH

If you use results from this analysis in published research, please cite:

Parker, K. C., M. A. Bednarek, and J. E. Coligan. 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152:163.

Return to BIMAS home page

Immunogenetics (1999) 50:213–219

REVIEW

Hans-Georg Rammensee · Jutta Bachmann
Niels Philipp Nikolaus Emmerich
Oskar Alexander Bachor · Stefan Stevanović

SYFPEITHI: database for MHC ligands and peptide motifs

Abstract The first version of the major histocompatibility complex (MHC) databank SYFPEITHI: database for MHC ligands and peptide motifs, is now available to the general public. It contains a collection of MHC class I and class II ligands and peptide motifs of humans and other species, such as apes, cattle, chicken, and mouse, for example, and is continuously updated. All motifs currently available are accessible as individual entries. Searches for MHC alleles, MHC motifs, natural ligands, T-cell epitopes, source proteins/organisms and references are possible. Hyperlinks to the EMBL and PubMed databases are included. In addition, ligand predictions are available for a number of MHC allelic products. The database content is restricted to published data only.

Key words MHC · Peptide motif · T-cell epitope

Introduction

The function of MHC molecules is the transfer of information about the current stock of proteins within a cell to the cell surface, thus enabling the immune system to react if necessary, for example by inducing cytotoxic T lymphocytes to kill virus-infected cells or by activating B cells by a helper T lymphocyte. In this context, the peptide specificity of MHC class I and class II molecules is important for the selection of relevant peptides.

The first collection of MHC ligands and peptide motifs was published as the First Listing in the anniversary issue of Immunogenetics in 1995 (Rammensee et al. 1995) and already contained at this time a few hundred entries, mainly consisting of human ligands and T-cell epitopes. By 1997, an enormous amount of information on the peptides associated with MHC molecules had been accumulated and an update of the first collection was published (Rammensee et al. 1997). This collection contained not only MHC peptide motifs, MHC ligands, and T-cell epitopes but also the amino acid sequences of MHC molecules, to enable the elucidation of the structural basis of MHC motifs by analyzing the nature of the pockets involved in the binding process. Since then, even more MHC motifs have become available. In order to meet the needs of colleagues involved with MHC-associated peptides and to provide a handy source of MHC ligands and epitopes and related entries, the database SYFPEITHI was designed and published via the World-Wide-Web (WWW). The name SYFPEITHI was chosen to acknowledge the first MHC-eluted peptide that was directly sequenced (Falk et al. 1991). Data related to peptides eluted from MHC molecules have been compiled in the database; at present, it comprises approximately 2000 entries reported for human, mouse, rat, ape, cattle, and chicken MHC alleles. All entries are linked to the respective sequences of the EMBL database and the abstracts published in PubMed. The possibility of epitope predictions for a limited number of MHC motifs has been opened, and in the near future the molecular mass of the peptides will complement the collection.

H.-G. Rammensee (✉) · J. Bachmann · N.P.N. Emmerich
S. Stevanović
Universität Tübingen, Interfakultäres Institut für Zellbiologie,
Abteilung Immunologie, Auf der Morgenstelle 15,
D-72076 Tübingen, Germany
e-mail: rammensee@uni-tuebingen.de
Tel.: +49-7071-2987628
Fax: +49-7071-295653

O.A. Bachor
EMS Medicinal and Scientific Dataprocessing,
Willy-Brandt-Platz 6, D-68161 Mannheim, Germany
EMS.Mannheim@t-online.de

Materials and methods

All data are stored centrally in a relational client-server database system (RDBMS). The main table of the RDBMS contains examples of ligands and T-cell epitopes, as well as additional informa-

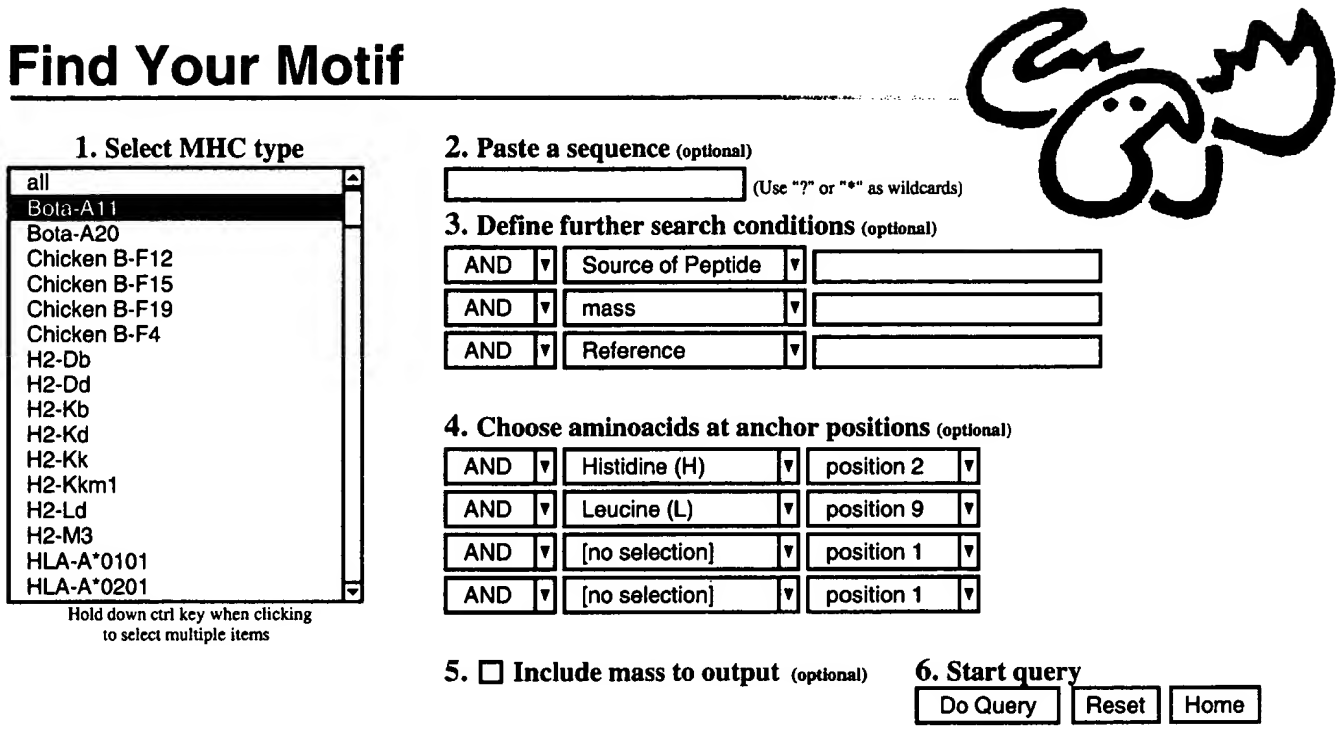


Fig. 1 'Find Your Motif' section of the SYFPEITHI database

tion on the specific role (anchor and auxiliary anchor amino acids) and position of each individual amino acid (aa). When data are browsed, the information is transformed to present a formatted version in which anchor aa are given in bold letters and auxiliary anchors are underlined. The main table is linked in a many-to-one relationship to the list of sources, and in a many-to-many relationship with the table of references. Each record in the table of source proteins refers to the specific EMBL ID, whereas each entry in the table of references is linked to the accession number (AN) of the reference in the NLM-PubMed database.

Table 1 Motif pattern for the prediction of HLA-B*1510 ligands

AA	1	2	3	4	5	6	7	8	9
A	0	0	1	0	0	0	0	1	0
C	0	0	0	0	0	0	0	0	0
D	0	0	0	1	0	0	0	0	0
E	1	0	1	1	0	0	0	1	0
F	0	0	0	0	0	0	0	0	**6**
G	1	0	0	1	1	0	0	0	0
H	0	**10**	0	0	0	0	0	0	0
I	2	0	0	0	0	1	0	0	0
K	0	0	0	1	0	1	0	0	0
L	0	0	0	0	0	0	0	0	**10**
M	0	0	0	0	0	1	0	0	**6**
N	0	0	0	0	1	0	0	0	0
P	0	0	0	2	1	1	1	0	0
Q	0	0	0	1	0	0	0	0	0
R	0	0	0	0	0	1	2	2	0
S	0	0	1	0	0	0	0	0	0
T	1	0	0	0	0	0	0	1	0
V	0	0	0	1	0	1	2	2	0
W	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0
Y	1	0	0	0	0	0	0	0	0

Database retrieval can be performed on any HTML-browser supporting JavaScript. The main page of the database (http://www.uni-tuebingen.de/uni/kxi/) offers three sections: "Find Your Motif", "Epitope prediction" and "Information". After a preselection of one or multiple MHC-types, the "Find Your Motif" section (Fig. 1) allows the user to search for a complete or truncated sequence of up to nine aa, a given peptide source, or a reference. The search can be narrowed down even further by choosing a specific aa on a given position as anchor or auxiliary anchor. All search criteria may also be combined to obtain a complex analysis. When a search is performed, an SQL-query is generated and the results are presented on a dynamically composed HTML page. The page of results lists the MHC type, motifs, peptide sources, and references. From each peptide source and each reference, a hyperlink to the EMBL or PubMed database is generated, respectively.

The algorithm used for epitope prediction is written in Object-Pascal. In brief, a two-dimensional data array is built up, where the letters of the aa represent the row index and the pocket numbers represent the column index (Table 1). The scores in the array-cells of the matrix shown in Table 1 can be addressed directly by a pair of indices. Starting at the first aa, the sequence is then divided into octa-, nona- or decamers and for each oligomer the sum of the scores of the aa contained is calculated. The process is then repeated until the end of the sequence is reached. Amino acids that frequently occur in anchor positions are given the value 10, the value 8 is given to amino acids present in a significant number of ligands, and 6 for rarely occurring residues; amino acids of auxiliary anchor positions are given the value 6, less frequent residues of the same set have a coefficient of 4; preferred amino acids have coefficients of 1–4 according to the strength of signals in pool sequencing or the occurrence in individual sequences. Amino acids that are regarded as unfavorable for binding have a coefficient of –1 to –3. These values are taken into account in the algorithm.

Results and discussion

The main contents of the first SYFPEITHI release are shown in Table 3. Since the First Listing, a nearly five-

Table 2 Peptide motif and natural ligands of HLA-DRB1*0301

	Position	Source protein	Reference
	1 2 3 4 5 6 7 8 9		
Anchors			Malcherek et al. 1993; Geluk et al. 1992, 1994
	L D K Y		
	I R L		
	F E F		
	M Q		
	V N		
Examples for ligands			
	I S N Q **L** T L **D** S N T **K** **Y** F H K L N	Apolipoprotein B-100 (2877–2894)	Malcherek et al. 1993
	I S N Q **L** T L **D** S N T **K** **Y** F H K L	Apolipoprotein B-100 (2877–2893)	Malcherek et al. 1993
	I S N Q **L** T L **D** S N T **K** **Y** F H K	Apolipoprotein B-100 (2877–2892)	Malcherek et al. 1993
	V D T **F** L E **D** V **K** N L **Y** H S E A	α1-Antitrypsion (149–164)	Malcherek et al. 1993
	K P R A I V V **D** P V H G **F** M Y	LDL receptor (518–532)	Malcherek et al. 1993
	K Q T **I** S P **D** **Y** **R** N M I	IG2a (384–395)	Malcherek et al. 1993
	Y P **D** **F** I M **D** P **K** E K D K V	Unknown	Malcherek et al. 1993
	N I Q **L** I N **D** Q E V A R F D	Unknown	Malcherek et al. 1993
	L L S **F** V R **D** L N Q Y R A D I	Transferrin receptor (618–632)	Malcherek et al. 1993
L P K P P K P V S K **M** **R** M A T P L		Invariant chain (97–111)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
L P K P P K P V S K **M** **R** M A T P L L M Q A L P		Invariant chain (97–119)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
L P K P P K P V S K **M** **R** M A T P L L M Q A L P M		Invariant chain (97–120)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
P K P P K P V S K **M** **R** M A T P L		Invariant chain (98–113)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
P K P P K P V S K **M** **R** M A T P L L M Q A		Invariant chain (98–117)	Riberdy et al. 1992; Chicz et al. 1993:Sette et al. 1992
K P P K P V S K **M** **R** M A T P L L M Q		Invariant chain (99–116)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
K P P K P V S K **M** **R** M A T P L L M Q A L P M		Invariant chain (99–119)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
V D D T Q F **V** **R** F **D** S D A A S Q		HLA-A30 (52–67)	Chicz et al. 1993
A T K Y G N **M** **T** E **D** H V M H L L Q N A		Invariant chain (131–149)	Chicz et al. 1993
V F L L L L A **D** K V P E T S L S		ACh receptor (289–304)	Chicz et al. 1993
L N K **I** L L **D** E **Q** A Q W K		ICAM-2 (64–76)	Chicz et al. 1993
G P P K L **D** **I** R K E E **K** **Q** **I** M I D I F H		IFNγ receptor (128–147)	Chicz et al. 1993
G P P K L **D** **I** R K E E **K** **Q** **I** M I D I F H P		IFNγ receptor (128–148)	Chicz et al. 1993
G F K **A** **I** R P **D** K **K** S N P I I R T V		Cyt-B5 (155–172)	Chicz et al. 1993
Y A N **I** L L **D** **R** **R** V P Q T D M T F		Apolipoprotein B-100 (1207–1224)	Chicz et al. 1993
N L F **L** K S **D** G **R** **I** K Y T L N K N S L K		Apolipoprotein B-100 (1276–1295)	Chicz et al. 1993
I P D N L F **L** K S **D** G **R** **I** K Y T L N K N		Apolipoprotein B-100 (1273–1292)	Chicz et al. 1993
I P D N L F **L** K S **D** G **R** **I** K Y T L N K		Apolipoprotein B-100 (1273–1291)	Chicz et al. 1993
I P D N L F **L** K S **D** G **R** **I** K Y T L N		Apolipoprotein B-100 (1273–1290)	Malcherek et al. 1993; Chicz et al. 1993
I P D N L F **L** K S **D** G **R** **I** K Y T L		Apolipoprotein B-100 (1273–1289)	Chicz et al. 1993
N L F **L** K S **D** G **R** **I** K Y T L N K		Apolipoprotein B-100 (1276–1291)	Chicz et al. 1993
N L F **L** K S **D** G **R** **I** K Y T L N		Apolipoprotein B-100 (1276–1290)	Chicz et al. 1993
V T T **L** N S **D** L **K** Y N A L D L T N		Apolipoprotein B-100 (1294–1310)	Chicz et al. 1993
V G S **D** **W** **R** F L R G Y H Q Y A		HLA-A2 (103–117)	Chicz et al. 1993
T-cell epitopes			
G D V V A V V **D** **I** K E K G K D K W I E L K		Lol pol. P1 (171–190)	Geluk et al. 1994
K T **I** A Y **D** E E A R R		HSP65 (cattle) (3–13)	Hawes et al. 1995
M G R S I **K** V Q **L** Q		M. tuberculosis 30/31 kD protein (56–65)	Geluk et al. 1997
S D K N P **L** F L **D** E Q L I		M. tuberculosis HSP70 (257–269)	Geluk et al. 1997

fold increase in MHC motifs has been registered; the database SYFPEITHI now includes approximately 200 peptide motifs and 2000 peptide sequences. Each entry contains the peptide sequence, its MHC specificity, source protein, anchor positions, and publication references with links to the respective sequences in the EMBL databank and to the NLM-literature database PubMed. We refrained from including sequences other than those confirmed to ensure that only natural ligands and T-cell epitopes that are relevant for the respective MHC molecule are listed. A different database, MHCPEP (http://wehih.wehi.edu.au/mhcpep/), offered at the WEHI in Australia (Brusić et al. 1998) includes as many as 13 000 entries, in which submissions of preliminary data are included. MHCPEP also contains peptides that have been reported to bind to MHC in the absence of any functional data. Such peptides have been omitted from the SYFPEITHI database.

Table 4 depicts the MHC ligands of HLA-B*1510 with the anchoring amino acids printed in bold letters as an example of a database entry. If available, auxiliary anchors are underlined, and preferred residues are also given. The typical length of a class I ligand comprises 9 amino acids. Below the anchor positions a list of ligands and T-cell epitopes specific for the respective molecule follows. This includes the respective protein sources and references, which are directly linked to other databases available on-line. Every single MHC allele has its individual peptide specificity that is defined primarily by the position and specificity of the pockets that accommodate the side chains of the anchoring amino acids and in the second place by interactions of non-anchoring amino acid residues of the peptides.

Class II ligands consist of 12 to 25 amino acids, nine of which occupy the binding groove; between two and four are anchored in the pockets. As in the case of class

Table 3 MHC molecules currently included in the database

Class I

Human:

HLA-A1
HLA-A*0201
HLA-A*0202
HLA-A*0203
HLA-A*0204
HLA-A*0205
HLA-A*0206
HLA-A*0207
HLA-A*0209
HLA-A*0214
HLA-A3
HLA-A*0301
HLA-A*1101
HLA-A24
HLA-A*2402
HLA-A25
HLA-A26
HLA-A*2601
HLA-A*2602
HLA-A*2603
HLA-A29
HLA-A*2902
HLA-A*3001
HLA-A*3002
HLA-A*3003
HLA-A*3004
HLA-A*3101
HLA-A32
HLA-A*3301
HLA-A*3302
HLA-A*6601
HLA-A*6801
HLA-A*6802
HLA-A*6901
HLA-A*7401
HLA-B7
HLA-B*0702
HLA-B*0703
HLA-B*0705
HLA-B8
HLA-B*0801
HLA-B*0802
HLA-B13
HLA-B14
HLA-B15
HLA-B*1501 (B62)
HLA-B*1502
HLA-B*1503
HLA-B*1508
HLA-B*1509
HLA-B*1510
HLA-B*1513
HLA-B*1516
HLA-B*1517
HLA-B17
HLA-B18
HLA-B*1801
HLA-B22
HLA-B27
HLA-B*2702
HLA-B*2703
HLA-B*2704
HLA-B*2705
HLA-B*2706
HLA-B*2707
HLA-B*2709
HLA-B35
HLA-B*3501
HLA-B*3503
HLA-B37
HLA-B*3701
HLA-B*3801
HLA-B39
HLA-B*39011
HLA-B*3902
HLA-B*40011
HLA-B*40012 (B60)
HLA-B*4002
HLA-B*4006 (B61)
HLA-B42
HLA-B44
HLA-B*4402
HLA-B*4403
HLA-B*4405
HLA-B45
HLA-B*4501
HLA-B*4601
HLA-B*4801
HLA-B51
HLA-B*5101
HLA-B*5102
HLA-B*5103
HLA-B52
HLA-B*5201
HLA-B*5301
HLA-B*5401
HLA-B*5501
HLA-B*5502
HLA-B*5601
HLA-B57
HLA-B*5701
HLA-B*5702
HLA-B58
HLA-B*5801
HLA-B*5802
HLA-B*6701
HLA-B*7301
HLA-B*7801
HLA-Cw*0301
HLA-Cw*0304
HLA-Cw*0401
HLA-Cw*0601
HLA-Cw*0602
HLA-Cw*0702
HLA-Cw*1601
HLA-E
HLA-G

Macacca fascicularis:

Mafa-A2

Macacca mulatta:

Mamu-A*01
Mamu-A*08
Mamu-B*12

Pan paniscus:

Papa-Q*06

Pan troglodytes:

Patr-A*04
Patr-A*11
Patr-B*01
Patr-B*13
Patr-B*16

Saguinus oedipus:

Saoe-G*01
Saoe-G*06 or Saoe-G*04
Saoe-G*08
Saoe-G, unassigned

Mouse:

H2-D^b
H2-D^d
H2-K^b
H2-K^d
H2-K^k
H2-K^{km1}
H2-L^d
H2-M3
H2-M3f
Qa-1a
Qa-1b
Qa-2

Rat:

RT1.A^a
RT1.A^a (Tap2A-associated
RT1.A^a (Tap2B-associated)
RT1.A^u
RT1.$A1^c$
RT1.A^l

Cattle:

Bota-A11
Bota-A20

Chicken:

Chicken B-F12
Chicken B-F15
Chicken B-F19
Chicken B-F4

Pig:

D/d

Class II

Human:

HLA-DQ1
HLA-DQ6 or -DQ2
HLA-DQA1*0201/DQB1*0202
HLA-DQA1*0301/DQB1*0302 (DQ8)
HLA-DQA1*0301/DQB1*0303 (DQ9)
HLA-DQA1*0501/DQB1*0201(DQ2)
HLA-DQA1*0501/DQB1*0301(DQ7)
HLA-DQB1*0201
HLA-DQA1*0301/DQB1*0301 (DQ3.1)
HLA-DQA1*0301/DQB1*0401 (DQ4)
HLA-DQB1*0603
HLA-DR2
HLA-DR2 (DRB5*0101 or DRB1*1501)
HLA-DRB1*0101
HLA-DRB1*0102
HLA-DR17 or DRw52
HLA-DR3
HLA-DRB1*0301 (DR17)
HLA-DRB1*0401 (DR4Dw4)
HLA-DRB1*0402 (DR4Dw10)
HLA-DRB1*0403 (DR3Dw13)
HLA-DRB1*0404 (DR4Dw14)
HLA-DRB1*0405 (DR4Dw15)
HLA-DRB1*0406
HLA-DRB1*0407
HLA-DR7
HLA-DRB1*0701
HLA-DR8
HLA-DRB1*0801
HLA-DRB1*0802
HLA-DRB1*08032
HLA-DRB1*0901
HLA-DRB1*1001
HLA-DR11
HLA-DR11 or Dw52
HLA-DRB1*1101
HLA-DRB1*1102
HLA-DRB1*1104
HLA-DRB1*1201
HLA-DRB1*1301
HLA-DRB1*1302
HLA-DRB1*1401
HLA-DRB1*1405
HLA-DRB1*1501 (DR2b)
HLA-DRB1*1502
HLA-DRB1*1601
HLA-DRB1, unassigned
HLA-DRB3*0101
HLA-DRB3*0202 (DR52Dw25)
HLA-DRB3*0301 (DR52Dw26)
HLA-DRB4
HLA-DRB4*0101
HLA-DRB5*0101 (DR2a)
HLA-DRB5*0202
HLA-DRB, unassigned
HLA-DRw11
HLA-DRw52
HLA-DRw52 or HLA-DQ2

Macacca mulatta:

Mamu-DRB1*0406
Mamu-DRB*W201

Mouse:

H2-A^b, H2-A^d, H2-A^{g7}, H2-A^k, H2-A^s, H2-A^u, H2-E^b, H2-E^d

Rat:

RT1.B^l

Table 4 Peptide motif and natural ligands of HLA-B*1510 (Seeger and co-workers, in press)

	Position 1 2 3 4 5 6 7 8 9	Source	Accession No. EMBL database
Anchor residues	**H L**		
Preferred residues	I E P A V V V F Y A D G I R R M T S E P M P E G G N K T E K R A Q V		
Examples for ligands	G H D P R A Q G T **L**	HLA-DP α chain (220–229)	X00457
	D H C V A H K **L**	Cytochrome C reductase (66–73)	M36647
	I **H** E D S T N R R R **L**	Heat shock protein 90 b (440–450)	M16660
	E **H** A H N M R V **M**	Elongation factor 2 (489–497)	M19997
	G H L E N N P A **L**	60 S acidic rib. protein PQ (67–75)	M17885
	H H S G A K V V **L**	Chaperonin cont. TCP-1 η (282–290)	AF026292
	I **H** D P G R G A P **L**	60 S ribosomal protein L8 (49–58)	Z28407
	T **H** T Q P G V Q **L**	Septin 2 homologue (70–78)	D50918
	T **H** Y V A P R R **L**	Transcription activator SNF2L4 (899–907)	U29175
	Y **H** G H G V S A **F**	Human EST	T96718
	Y Q E K G V R V **L**	Actin-related protein Arp2 (402–410)	AF006082
	I **H** E P E P H I **L**	Cyclin-dep. kin. reg. subunit 1 (59–67)	X54941
	A H S T I M P R **L**	DNA repl. lic. factor MCM4 (694–702)	X74794
	E **H** A G V I S V **L**	HBV X interacting protein (40–48)	AF029890

I ligands, the nonanchoring amino acids play a secondary, but still significant role. A number of examples of ligands specific for HLA-DRB1*0301 are given in Table 2.

Information on the allelic specificity of the motifs, including preferred and unfavored residues beyond the anchor, enable the prediction of MHC ligands and T-cell epitopes. The second section of the database is dedicated to "Epitope prediction". The main objective of performing theoretical predictions is simply to save time. Instead of synthesizing and testing dozens or even hundreds of peptides, a preselection of a small set of peptides is made. The sequence of the protein or its gene, the restriction element, and its respective motif have to be available.

Motif-based predictions result in a list of peptides that have a high probability of being presented by MHC class I molecules. The prediction method used in the database is based on the principle of a building-up pattern (see Material and methods).The pattern used for the prediction of HLA-B*1510 nonamers is shown in Table 1. The values of all possible nonamers of a given sequence are added together and the optimal T-cell epitope is expected within the ten high-scoring peptides of each protein. Scores for all source proteins of nonamer peptides and predicted epitopes specific for HLA-B*1510 are given in Table 5. The correct and therefore naturally presented T-cell epitope from any source protein is assumed to appear among the top 2% of peptides in the high score list in more than 90% of predictions. However, theoretical approaches cannot always guarantee success. About 10% of the predictions are still unable to identify the respective epitope because not all details of the motifs are entirely known. Epitope predictions are only available at present for a small number of MHC alleles, since all predictions undergo several rounds of thorough cross-checking and a certain number of natural ligands and/or T-cell epitopes have to be available to ensure an accurate and reliable prediction (Table 6). The occurrence of unusual anchor amino acids can be compensated by refining the motif, but still a small percentage of epitopes do not act in accordance with any rules of the respective motif. In these cases, the theoretical approach will fail and experimental epitope mapping is unavoidable. SYFPEITHI is the only freely available MHC database in the internet apart from an HLA peptide binding prediction offered by the NIH (http://www-bimas.dcrt.nih.gov/cgi-bin/molbio/hla_bind/). The NIH prediction relies upon binding data (Parker et al. 1994) and estimates the half-time of dissociation of a given MHC-peptide complex. EpiMer (http://www.brown.edu/Research/TB-HIVLab/Epimatrix/epimerlink.html; Meister et al. 1995; Roberts et al. 1996) is a commercial site.

Other programs for the prediction of T-cell epitopes are distributed on disk (e.g., Motifs, D'Amaro et al. 1995; Davenport et al. 1995; Devereux et al. 1984; Fleckenstein et al. 1997; Hammer et al. 1994). Prediction patterns for class II are not yet included in our database due to the highly degenerate anchor positions in most MHC class II motifs. These degenerate anchor positions render the prediction of MHC class II ligands rather difficult. If all possible anchor residues of an MHC class II peptide motif are included within a search pattern, a very high number of possible natural MHC II ligands will be predicted. Only few of the class II peptide motifs encompass anchor residues specific

Table 5 Motif prediction of HLA-B*1510 self peptides and epitopes

Actin-related protein 2 (ARP2; human)

AA pos.	Score	Sequence
378	19	YQEKGVRVL
227	15	IEQEQKLAL
29	14	EHIFPALVG
60	14	EASELRSML
164	14	THICPVYEG

Septin 2 homologue (SEP2; human)

AA pos.	Score	Sequence
70	25	THTQPGVQL
156	22	GHSLKSLDL
371	22	LHQDEKKKL
129	17	EELKIRRVL
84	15	DLQESNVRL

60 S acidic ribosomal protein RQ (RLA0; human)

AA pos.	Score	Sequence
67	25	GHLENNPAL
80	19	PHIRGNVGF
241	18	IINGYKRVL
46	15	SLRGKAVVL
196	15	GSIYNPEVL

Elongation factor 2 (EF2; human)

AA pos.	Score	Sequence
489	24	EHAHNMRVM
356	18	IHLPSPVTA
147	17	IAERIKPVL
491	17	AHNMRVMKF
843	16	GLKEGIPAL

Transcription activator SNF2L4 (SN24; human)

AA pos.	Score	Sequence
899	27	THYVAPRRL
1009	24	RHMQAKGVL
620	23	IHVESGKIL
889	23	HHCKLTQVL
968	23	LHKVLRPFL
252	18	PHGMGGPNM
522	18	IEKERMRRL

Cyclin-dependent kinase regulatory subunit (CKS1; human)

AA pos.	Score	Sequence
59	26	IHEPEPHIL
16	14	EFEYRHVML
64	14	PHILLFRRP
29	13	AKLVPKTHL
38	13	MSESEWRNL

DNA replication license factor (MCM4; human)

AA pos.	Score	Sequence
694	25	AHSTIMPRL
331	21	CHTTHSMAL
255	20	EHQIQVRPF
440	16	LFSEKRVEL
741	16	AEAHAKVRL

enough to make motif-based predictions worthwhile, such as HLA-DRB1*0301, for example. In addition, anchor residues are often not used by class II ligands and epitopes. Inclusion of possibilities to predict class II ligands is in preparation.

Table 5 Continued

HIV1 GAG

AA pos.	Score	Sequence
192	22	GHQAAMQML
69	15	TGSEELRSL
143	15	HQAISPRTL
354	15	GPGHKARVL
42	14	RFAVNPGLL

Influenza A PR8/34 nucleoprotein

AA pos.	Score	Sequence
333	21	CHSAAFEDL
323	18	AHKSQLVWM
100	15	VNGKWMREL
336	15	AAFEDLRVL
48	14	KLSDYEGRL

Table 6 MHC motifs included in the database for epitope predicitions

Human	Mouse
HLA-A*0201	H2-D^b
HLA-A*0202	H2-K^b
HLA-A*0203	H2-K^d
HLA-A1	H2-K^k
HLA-A26	H2-L^d
HLA-B*0702	
HLA-B*1510	
HLA-B*2705	
HLA-B8	

Availability

SYFPEITHI is freely available online at the Interfakultäres Institut für Zellbiologie, Abteilung Immunologie, Tübingen (WWW site address: http://www.uni-tuebingen.de/uni/kxi).

Acknowledgments We acknowledge financial support of the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 510, Projekt E2) and the European Union and Biotech CT95-0263; Biomed CT95-1627. We also thank Vladimir Brusić for his valuable comments.

References

Brusić V, Rudy G, Kyne AP, Harrison LC (1998) MHCPEP, a database of MHC-binding peptides: update 1997. Nucleic Acids Res 26:368–371

Chicz RM, Urban RG, Gorga JC, Vignali DA, Lane WS, Strominger JL (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J Exp Med 178:27–47

D'Amaro J, Houbiers JG, Drijfhout JW, Brandt RM, Schipper R, Bavinck JN, Melief CJ, Kast WM (1995) A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide-binding motifs. Hum Immunol 43:13–18

Davenport MP, Ho Shon IAP, Hill AVS (1995) An empirical method for the prediction of T-cell epitopes. Immunogenetics 42:392–397

Devereux J, Haeberli P, Smithies OA (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290–296

Fleckenstein B, Jung G, Wiesmüller KH (1997) Prediction and Design of new MHC class II-ligands based on the Activity Pattern of a Synthetic Undecapeptide Library. In: Brown F, Burton D, Doherty P, Mekalanos J, Norrby E (eds) Vaccines 97: Molecular approaches to the control of infectious diseases, Cold Spring Harbor Laboratory Press, pp 65–70

Geluk A, van Meijgaarden KE, Southwood S, Oseroff C, Drijfhout JW, de Vries RR, Ottenhoff TH, Sette A (1994) HLA-DR3 molecules can bind peptides carrying two alternative specific submotifs. J Immunol 152:5742–5748

Geluk A, van Meijgaarden KE, de Vries RR, Sette A, Ottenhoff TH (1997) A DR17-restricted T cell epitope from a secreted Mycobacterium tuberculosis antigen only binds to DR17 molecules at neutral pH. Eur J Immunol 27:842–847

Geluk A, Van Meijgaarden KE, Janson AA, Drijfhout JW, Meloen RH, De Vries RR, Ottenhoff TH (1992) Functional analysis of DR17(DR3)-restricted mycobacterial T cell epitopes reveals DR17-binding motif and enables the design of allele-specific competitor peptides. J Immunol 149:2864–2871

Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F(1994) Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. J Exp Med 180:2353–2358

Hawes GE, Struyk L, Godthelp BC, van den Elsen PJ (1995) Limited restriction in the TCR-alpha beta V region usage of antigen-specific clones. Recognition of myelin basic protein (amino acids 84–102) and Mycobacterium bovis 65-kDa heat shock protein (amino acids 3–13) by T cell clones established from peripheral blood mononuclear cells of monozygotic twins and HLA-dientical individuals. J Immunol 154:555–566

Malcherek G, Falk K, Rötzschke O, Rammensee HG, Stevanovic S, Gnau V, Jung G, Melms A (1993) Natural peptide ligand motifs of two HLA molecules associated with myasthenia gravis. Int Immunol 5:1229–37

Meister GE, Roberts CGP, Berzofsky JA, De Groot AS (1995) Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from Mycobacterium tuberculosis and HIV protein sequences. Vaccine 13:581–91

Parker KC, Bednarek MA, Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152:163–175

Rammensee HG, Bachmann J, Stevanović S (1997) MHC ligands and peptide motifs. Landes Biosciences, Texas

Rammensee HG, Friede T, Stevanović S (1995) MHC ligands and peptide motifs: first listing. Immunogenetics 41:178–228

Riberdy JM, Newcomb JR, Surman MJ, Barbosa JA, Cresswell P (1992) HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. Nature 360:474–477

Roberts CG, Meister GE, Jesdale BM, Lieberman J, Berzofsky JA, De Groot AS (1996) Prediction of HIV peptide epitopes by a novel algorithm. AIDS Res Hum Retroviruses 12:593–610

Seeger FH, Schirle M, Keilholz W, Rammensee HG, Stevanović (in press) Peptide motif of HLA-B*1510. Immunogenetics 49

Sette A et al. (1992) Invariant chain peptides in most HLA-DR molecules of an antigen-processing mutant. Science 258:1801–1804

ELSEVIER

Journal of Immunological Methods 257 (2001) 1–16

JIM
Journal of Immunological Methods

www.elsevier.com/locate/jim

Review

Combining computer algorithms with experimental approaches permits the rapid and accurate identification of T cell epitopes from defined antigens

Markus Schirle, Toni Weinschenk, Stefan Stevanović *

Department of Immunology, Institute for Cell Biology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany

Received 19 April 2001; accepted 2 July 2001

Abstract

The identification of T cell epitopes from immunologically relevant antigens remains a critical step in the development of vaccines and methods for monitoring of T cell responses. This review presents an overview of strategies that employ computer algorithms for the selection of candidate peptides from defined proteins and subsequent verification of their in vivo relevance by experimental approaches.

Several computer algorithms are currently being used for epitope prediction of various major histocompatibility complex (MHC) class I and II molecules, based either on the analysis of natural MHC ligands or on the binding properties of synthetic peptides. Moreover, the analysis of proteasomal digests of peptides and whole proteins has led to the development of algorithms for the prediction of proteasomal cleavages. In order to verify the generation of the predicted peptides during antigen processing in vivo as well as their immunogenic potential, several experimental approaches have been pursued in the recent past. Mass spectrometry-based bioanalytical approaches have been used specifically to detect predicted peptides among isolated natural ligands. Other strategies employ various methods for the stimulation of primary T cell responses against the predicted peptides and subsequent testing of the recognition pattern towards target cells that express the antigen.

Keywords: T cell; Epitope prediction; HLA; Peptide; Computer algorithm

1. Introduction

Major histocompatibility complex (MHC) molecules are highly polymorphic cell surface molecules that present peptidic ligands to cells of the T cell compartment of the immune system. MHC class I ligands have a typical length of 8–12 amino acids and are derived from endogenously expressed proteins that are degraded by cytosolic proteases, most notably the proteasome. The proteolytic fragments are transported into the endoplasmic reticulum in an

Abbreviations: ANN, artificial neural network; CTL, cytotoxic T lymphocyte; ELISA, enzyme-linked immunosorbent assay; FACS, fluorescence-activated cell sorting; HLA, human leukocyte antigen; MHC, major histocompatibility complex; MSMS, tandem mass spectrometry; RT-PCR, reverse transcriptase-polymerase chain reaction; TAA, tumor-associated antigen; TAP, transporter associated with antigen processing; TCR, T cell receptor; WWW, world wide web.

* Corresponding author. Tel.: +49-7071-2987645; fax: +49-7071-295653.

E-mail address: stefan.stevanovic@uni-tuebingen.de (S. Stevanović).

0022-1759/01/$ - see front matter
PII: S0022-1759(01)00459-8

ATP-dependent fashion by the transporter associated with antigen processing (TAP), where they bind to newly synthesized empty MHC class I molecules. The MHC-peptide complex is subsequently transported to the cell surface and can be recognized by the T cell receptor (TCR) of $CD8^+$ cytotoxic T cells (CTL). MHC class II ligands have a more variable length of 9–25 amino acids and are derived mainly from exogenous or transmembrane proteins, but also from cytosolic proteins that are degraded by various proteases which originate from the lysosomal compartment. After binding to MHC class II molecules in the late endosomal/lysosomal compartment, the MHC-peptide complex is also transported to the cell surface, where it can be recognized by the TCR of $CD4^+$ T cells.

Recognition of a peptide derived from a disease-associated protein, e.g., a viral or a tumor-specific protein, in the presence of a costimulatory signal leads to T cell activation and triggers a T cell-mediated immune response. Therefore, which peptide fragments of immunologically relevant antigens are available in the context of a certain MHC-molecule for recognition by T cells is crucial for the development of peptide-based or other defined antigen-based vaccines. Moreover, tools for monitoring specific T cell responses, such as MHC-peptide tetramers, intracellular cytokine staining or ELISPOT assays, depend on the identification of the relevant T cell epitope.

A major breakthrough in the identification of T cell epitopes was the discovery that ligands of a certain MHC-molecule carry chemically related amino acids in certain positions of their primary sequence, which permits the definition of a "peptide motif" for every MHC allele (Falk et al., 1991). This information was rapidly used to predict potential epitopes from protein sequences (Pamer et al., 1991; Rötzschke et al., 1991) and was the start signal for the so-called "reverse immunology" approach (Celis et al., 1994), which has since become the most successful strategy for the identification of T cell epitopes.

The first step of such studies is usually the computer-based prediction of potential MHC ligands from a protein of interest, which is followed by experiments to verify the natural processing of the predicted peptides and their recognition by T cells. The classical and most straightforward approach relies on the stimulation of primary T cell responses against the predicted peptides in vitro or in (transgenic) animals and subsequent testing of the recognition pattern of the generated T cells towards target cells that endogenously express the antigen. Alternatively, patient-derived T cells, which might have been already primed in vivo, are used. This "pure" reverse immunology approach has been used extensively in the recent past for the identification of CTL and, to a lesser extent, helper T cell epitopes from a wide variety of immunologically relevant proteins. However, a major drawback of this labor- and resource-intensive approach is the high failure rate, even if the peptide in question binds well to the MHC molecule and the generated T cells recognize target cells loaded with the synthetic peptide. The reason for this is either that the peptide is not produced by the processing machinery of the cell, or the T cells raised by primary peptide stimulation in vitro are of low affinity, or a combination of both.

In consequence, various groups have made attempts to decrease the number of candidate peptides by applying additional criteria to epitope prediction. For example, the increasing knowledge about specificities of other cellular components of the antigen processing machinery, most notably the proteasome and TAP, has made it possible to apply these specificities as additional filters to the selection of candidate peptides. Moreover, experimental selection criteria, such as MHC-peptide binding assays and the bioanalytical verification of presentation of predicted MHC-ligands, have been applied successfully.

This review tries to give a comprehensive overview of available computer-based prediction algorithms and strategies that combine the predictive power of these theoretical approaches with experimental approaches for a reliable and rapid identification of T cell epitopes.

Since it has been written from an immunologist's point of view, this review focuses mostly on prediction programs that have already been applied to the identification of novel epitopes. Furthermore, it should be noted that there are tremendous differences between different prediction softwares in terms of availability. Some programs are easily accessible for free on the world wide web (WWW); others are available from the authors of the corresponding pub-

lications; still others are available on a commercial basis only.

2. Computational predictions

2.1. Epitope prediction

The recognition of MHC-peptide complexes by the T cell receptor (TCR) is the first step in a series of cellular processes that lead to the initiation of T cell-mediated immunity. It is, therefore, not surprising that as soon as the peptidic nature of MHC ligands was discovered and before the rules for the binding of peptides to MHC molecules were elucidated, first attempts for the prediction of T cell antigenic structures from protein sequences were published. These predictions were based solely on comparisons of precursor peptide sequences known to contain T cell epitopes. The suggested characteristics of T cell epitopes were predicted secondary structures (amphipathic helices) (DeLisi and Berzofsky, 1985) or short primary sequence motifs (Rothbard and Taylor, 1988; Sette et al., 1989). It was the discovery of allele-specific motifs shared by eluted natural MHC ligands (Falk et al., 1991) that finally allowed the exact prediction of peptides from a given protein sequence presented on MHC class I molecules (Pamer et al., 1991; Rötzschke et al., 1991).

For MHC class II ligands, the elucidation of peptide motifs turned out to be more difficult due to the variable length of the ligands as well as the more degenerate anchor positions, which are sometimes not used at all. Several methods for the determination of MHC class II peptide motifs have been based on sequence information of natural ligands, including the alignment of natural ligands, pool sequence analysis, or substitution variants of known ligands (Rudensky et al., 1992; Chicz et al., 1992; Hunt et al., 1992a,b; Falk et al., 1994; Leighton et al., 1991). In 1992, the first HLA-DR-motif based on binding assays using a phage peptide display library was published (Hammer et al., 1994a,b).

In the beginning of the 1990s, the selection of candidate peptides was done manually or using simple pattern search programs by scanning protein sequences for main anchor positions only. However, the constant refinement of peptide motifs and knowledge about secondary anchor amino acids, disfavored amino acids and other chemical interactions important for binding to the MHC molecule quickly led to the development of more complex computational predictions.

Depending on the focus of research, the determination of peptide motifs has been approached from two sides.

On the one hand, the binding properties of particular MHC molecules have been analyzed by determining the binding of synthetic peptides from protein antigens using MHC-peptide binding assays (Ruppert et al., 1993; Kubo et al., 1994; Parker et al., 1994; Hammer et al., 1994a,b). The most refined descriptions of such "peptide-binding motifs" have been achieved by using positional scanning with random synthetic peptide libraries (Stryhn et al., 1996; Fleckenstein et al., 1996; Udaka et al., 2000).

On the other hand, the analysis of natural ligands by individual ligand analysis as characterized by Edman degradation (Falk et al., 1991; DiBrino et al., 1993; Corr et al., 1993; Falk et al., 1994) and tandem mass spectrometry (MSMS) (Hunt et al., 1992a,b) as well as Edman pool sequencing (Falk et al., 1991) leads to so-called "natural ligand motifs" that give information about the natural peptide repertoire presented by a certain MHC molecule. These motifs are, therefore, not only a result of the binding properties of this MHC molecule, but reflect also—to a certain extent—features of the antigen processing machinery, e.g., cleavage specificities of the proteasome or sequence requirements for transport into the ER by TAP. Indeed, MHC-binding motifs derived from random peptide libraries show significant differences, especially for the C-terminal amino acids, compared to motifs derived from natural ligands (Davenport et al., 1997). This raises the question of the in vivo relevance of motifs based on the binding properties of synthetic peptides.

2.1.1. MHC class I

Historically, MHC class I epitope prediction had a headstart over its MHC class II counterpart. The reasons for this include the defined length of MHC class I ligands of mostly nine amino acids and the more clearcut rules of MHC class I motifs that have been more clearly characterized. Currently, there are

two predictive algorithms available on the world wide web that allow unrestricted predictions free of charge: "BIMAS", developed by Parker et al. (1994) (http://www-bimas.dcrt.nih.gov/molbio/hla_bind/) and "SYFPEITHI", developed by our group (Rammensee et al., 1999) (http://www.syfpeithi.de). SYFPEITHI uses motif matrices deduced from refined motifs based on the pool sequence and single peptide analysis exclusively of natural ligands. Potential binders for various human, mouse and rat MHC class I molecules are ranked according to the presence of primary and secondary anchor amino acids as well as favored and disfavored amino acids.

BIMAS ranks potential human leukocyte antigen (HLA) binders according to predicted half-time dissociation of MHC-peptide complexes. While the motif matrix for HLA-A2 is based solely on binding studies, other matrices have been deduced from reported motifs, regardless of the method used for their identification. This might explain the similar results that were obtained with both algorithms in a recent study that identified HLA-B7-presented epitopes from the tumor-associated antigen (TAA) carcinoembryonic antigen (CEA) (Lu and Celis, 2000) since for HLA-B7 the reported motifs are based on natural ligands. Both algorithms have been used successfully during recent years for the identification of CTL epitopes from a wide range of antigens, e.g., from the human TAAs telomerase (BIMAS) (Vonderheide et al., 1999) and TRP2 (SYFPEITHI) (Sun et al., 2000).

Other prediction programs that have been used for the identification of MHC class I-presented epitopes are distributed on disk (e.g., "MOTIFS" (D'Amaro et al., 1995)), part of commercial websites [e.g., "EPIMATRIX" (De Groot et al., 1997) (http://www.brown.edu/Research/TB-HIV_Lab/epimatrix/epimatrix.html)] or are currently not publicly available (Sette et al., 1994).

The motif-based algorithms so far described are still relatively simple programs that consider every amino acid position in a peptide individually. For every position, a coefficient is assigned to each amino acid and the overall score or binding value of the peptide are determined by addition or multiplication of amino acid coefficients. Moreover, there are several reports employing artificial neural networks (ANN) (Brusic et al., 1994; Adams and Koziol, 1995; Gulukota et al., 1997; Milik et al., 1998) or the polynomial method (Gulukota et al., 1997). These approaches open up the possibility to consider amino acid preferences that depend on the properties of amino acids in other positions of the peptide.

Still another approach is based on structural information obtained by X-ray crystallography of peptide–MHC complexes (Altuvia et al., 1997; Schueler-Furman et al., 2000). While none of these latter, non motif-based approaches has actually been used to date for the identification of novel CTL epitopes, they could prove useful in the future.

Table 1 demonstrates the similarities and differences of some of the currently used MHC class I

Table 1
Comparison of MHC class I predictions
Epitopes identified using different epitope prediction programs were recalculated using programs that are available free of charge on interactive websites: the epitope prediction programs BIMAS and SYFPEITHI and a program for prediction of proteasomal cleavage sites PAPROC.

Sequence	Source (position)	HLA molecule	Program used for identification	Reference	SYFPEITHI rank/score	BIMAS rank/value	PAPROC *
ILAKFLHWL	human TERT (540–548)	A*0201	BIMAS	(Vonderheide et al., 1999)	1/30	1/1745.714	No
KVLEYVIKV	human MAGE-A1 (278–286)	A*0201	SYFPEITHI	(Pascolo et al., 2001)	3/26	1/743.189	Yes
ILHNGAYSL	human Her-2/neu (435–443)	A*0201	Sette et al.	(Rongcun et al., 1999)	3/27	16/36.316	Yes
HLSTAFARV	human G250 (254–262)	A*0201	MOTIFS	(Vissers et al., 1999)	19/22	38/4.493	No

* Prediction of proteasomal generation of C-terminus by human proteasome wild-type III.

prediction programs. It shows selected CTL epitopes that were identified using four different computer programs and also lists the corresponding results from SYFPEITHI, BIMAS and a prediction algorithm for proteasomal cleavage, "PAPROC" (see below), i.e., programs that are available at interactive websites at no cost. However, it should be noted that some studies do not give the results of the original prediction procedure, such as ranks or values, but only lists of peptides that were selected after performing additional tests.

2.1.2. MHC class II

The main obstacle for the prediction of MHC class II ligands has been the differing degree of degeneration of motifs. While some alleles, e.g., HLA-DRB1* 0405, show a strong preference for certain related amino acids in the anchor positions that is comparable to MHC class I motifs (in this case D, E in P9) (Rammensee et al., 1999), other alleles make the definition of primary anchor amino acids virtually impossible (e.g., for HLA-DRB1* 0401 all amino acids except G, P, F, W have been assigned anchor amino acids in P9) (Rammensee et al., 1999). To date, the focus has been on predictions of HLA-DR ligands, with only one report on predictions of HLA-DQ-restricted epitopes (Godkin et al., 1998). The first matrix-based prediction algorithms for HLA-DRB1* 0401 were based on the results of side-chain scanning experiments using simplified, polyalanine-based peptide libraries (Hammer et al., 1994a,b; Marshall et al., 1995). The work by Hammer and colleagues resulted in the commercial program "TEPITOPE", which focuses on promiscuous HLA-DR binding peptides, but also allows allele-specific predictions. It has been used in several studies, e.g., in the identification of an HLA-DR11-presented epitope from MAGE-A3 (Manici et al., 1999).

Another matrix-based HLA-DRB1* 0401-prediction algorithm, also based on binding studies using synthetic peptides (Southwood et al., 1998) has been used for the retrospective calculation of an epitope from human gp100 (Touloukian et al., 2000) that had been identified previously (Halder et al., 1997). This program is available free on request from the Surgery Branch of the NCI.

Still another customized computer algorithm also based on published HLA-DR binding motifs has been used for the identification of HLA-DR-presented promiscuous T cell epitopes from *Plasmodium falciparum* (Doolan et al., 2000). The first program for prediction of MHC class II based on natural ligands of two HLA-DR13 alleles was the PAP program (Davenport et al., 1995). This program, which can be obtained free from the author, uses corrected amino acid yields from Edman pool sequencing directly as coefficients in motif matrices. One of the following publications also reports for the first time predictions for HLA-DQ2 and -DQ8 and the identification of a novel HLA-DQ2-presented epitope from gliadin, the antigen of celiac disease (Godkin et al., 1998).

Currently, SYFPEITHI is the only interactive website offering unrestricted MHC class II predictions free of charge. As for MHC class I, the available predictions for HLA-DR and several mouse MHC class II alleles are based on motif matrices deduced from natural ligands by single ligand analysis and Edman pool sequencing.

For HLA-DRB1* 0401, predictions by artificial neural networks have already been applied successfully to the identification of helper T cell epitopes. Zarour and colleagues have used an ANN, based on the work of Honeyman et al. (1998) and Brusic et al. (1998), for the identification of epitopes from the TAAs Melan-A/MART-1 (Zarour et al., 2000a,b) and NY-ESO-1 (Zarour et al., 2000a,b). These prediction programs are currently not publicly available.

Finally, several publications have explored the possibilities of other more complex computational predictions. For example, Mallios (1999) has described an iterative algorithm to align and optimize an MHC class II-binding matrix based on sequences of known class II ligands alone or in combination with suggested motifs. Up to now, none of these predictions have been applied to the identification of T cell epitopes.

Table 2 shows examples of helper T cell epitopes that were identified by four different computer programs and also lists the corresponding results from SYFPEITHI and the HLA-DRB1* 0401 prediction program from the NCI. However, in contrast to MHC class I, comparisons of MHC class II predictions can be even more difficult for several reasons.

Table 2
Comparison of MHC class II predictions
Epitopes identified using different epitope prediction programs were recalculated using programs that are available free of charge either on an interactive website (SYFPEITHI) or by email (NCI).

Epitope sequence	Source (position)	HLA molecule	Reference	Program used for identification	SYFPEITHI rank/source	NCI rank/value
WNRQLYPEWTEAQRLD	human gp100 (44–59)	DRB1*0401	(Touloukian et al., 2000)	NCI	2/26	1/0.0195
RNGYRALMDKSLHVGTQCALTRR	human MART-1 (51–73)	DRB1*0401	(Zarour et al., 2000)	ANN	1/22	–/–
TSYVKVLHHMVKISG -----------	human MAGE-A3 (281–295)	DRB1*1101	(Manici et al., 1999)	TEPITOPE	8/22**	n.a.

Putative core sequences are underlined; ** best ranked peptides 284–298 (score 28) and 280–294 (score 24) also contain other putative core sequences (- - -); n.a., not available.

First, due to the variable length of MHC class II ligands, the identified epitopes may harbor more than one putative core sequence. Furthermore, some studies focus on promiscuous MHC class II-binding peptides, while others use allele-specific predictions.

2.2. Predictions of antigen processing

Much of the work combining computational predictions and applied immunology has focussed on MHC-peptide binding only. However, our increasing knowledge about the specificities of other cellular components involved in antigen processing, such as the proteasome and TAP, has been used to create comparable prediction algorithms with the future prospect of fusing them to epitope prediction programs.

The proteasome is a cytosolic multi-subunit protease that is, among other functions, involved in the generation of peptide ligands for MHC class I molecules and responsible for the generation of the correct C-terminus of CTL epitopes (Craiu et al., 1997; Stoltze et al., 1998). The proteolytically active core complex (20S proteasome) can associate with the 19S cap complex to form the 26S proteasome, which is responsible for the degradation of ubiquitin-marked proteins. The three different active sites show different proteolytic specificities, cleaving after large, hydrophobic amino acids, basic amino acids and acidic amino acids, respectively. Upon exposure of cells to interferon γ (IFNγ), the catalytically active subunits are replaced by three IFNγ-inducible subunits. This results in a shift in the overall cleavage specificity of this immunoproteasome towards hydrophobic amino acids, while cleavages after acidic amino acids are disfavored. This corresponds to the fact that almost all reported peptide motifs of vertebrate MHC molecules demand hydrophobic or basic C-terminal anchor amino acids.

An experimental approach was chosen by Kessler et al. (2001) for the identification of HLA-A* 0201-presented CTL epitopes from the TAA PRAME by an extended reverse immunology strategy that incorporates proteasomal digest analysis. Predicted HLA-A* 0201-binding peptides were tested in binding and stability assays, followed by an analysis of in vitro proteasome-mediated digestions of 27mer precursor peptides encompassing candidate epitopes. For only 4 out of 19 high affinity binders, efficient generation of the correct C-terminus was observed and all four peptides were found to be efficient CTL epitopes.

The first published prediction algorithm for cleavages by 20S proteasomes, "FRAGPREDICT", is based on published peptide cleavage data (Holzhütter et al., 1999; http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html). The PAPROC-algorithm (Nussbaum et al., 2001) http://www.paproc.de), developed by our group, is based on an evolutionary algorithm trained on cleavage data of 20S proteasomal digests of a whole protein substrate [enolase (Nussbaum et al., 1998)], and additional peptide digest data from the literature. While the currently available prediction programs are based only on semiquantitative data from human or yeast constitutive 20S proteasomal digests, more accurate predictions based on fully quantified protein digestion data as well as immunoproteasomal digestion data (Toes et al., 2001) are currently being developed. Another program, "NETCHOP" (Kesmir et al., manuscript submitted; http://www.cbs.dtu.dk/services/NetChop/) is an ANN that was trained on data from the analysis of residues at the termini and flanking regions of known natural MHC class I ligands. All these programs for the prediction of proteasomal cleavages are available free of charge on interactive websites.

The overall goal of combining proteasomal prediction algorithms with epitope prediction is already possible using combinations of FRAGPREDICT and PAPROC with SYFPEITHI and BIMAS at the website of the Max-Planck-Institute for Infection Biology (www.mpiib-berlin.mpg.de/MAPPP/). It should be noted that this program also predicts C-terminally extended precursor peptides.

An ANN-based device has also been reported for the prediction of human TAP peptide-binding affinities using a polyalanine-based peptide library (Daniel et al., 1998). The program predicted a higher affinity to TAP for eluted natural ligands compared to randomly selected peptides with similar binding affinities to the same HLA molecule, indicating that such predictions might be—at least for some HLA-alleles—a valuable addition for epitope prediction tools. However, this would probably require a more profound investigation of peptide translocation by TAP,

not only binding of peptides. The reported preferences of TAP for peptide translocations (Momburg et al., 1994) have not yet led to a prediction program.

3. Experimental verification of predicted epitopes

3.1. Analysis of binding properties

In most cases, epitope prediction is followed by binding studies with synthetic peptides in order to reduce the number of potential candidate peptides by discarding nonbinding peptides for the final analysis of recognition by peptide-specific T cells.

Binding of synthetic peptides to MHC molecules is mainly studied by assays based on competition (Sette et al., 1994; Olsen et al., 1994; Vogt et al., 1994), reconstitution of MHC-peptide complexes on the cell surface (Stuber et al., 1994; Zeh et al., 1994; Storkus et al., 1993), or stabilization of MHC-peptide complexes on TAP-deficient cells (Stuber et al., 1992).

In competition assays, the studied peptides compete with a labeled reference peptide of intermediate affinity for binding to MHC molecules. Such assays are performed either using whole cells or purified MHC molecules in solution. In both cases, the amount of labeled MHC-bound peptide is measured. For MHC class II-binding assays usually fluorescence dyes, like fluorescein or 7-amino-4-methylcoumarin-3-acetic acid (AMCA), are used for labeling the N-terminal amino acid of the MHC ligand. For MHC class I-binding studies, amino acid side chains have to be labeled by fluorescent dyes or by radioactive isotopes, such as ^{125}I without interfering with the more restricted binding properties of MHC class I molecules. Performing in-solution assays using purified MHC molecules, MHC-peptide complexes have to be separated from unbound free peptide by gel filtration (Sette et al., 1994), spun columns (Stryhn et al., 1997) or high-performance size exclusion chromatography (HPSEC) (Vogt et al., 1994). In addition, enzyme linked immunosorbent assay (ELISA)-based assays (Jensen, 1991; Hammer et al., 1995) or the more sensitive time-resolved fluorimetry in a europium fluoroimmunoassay (Tompkins et al., 1993) have been used as the readout. Competition-based binding assays result in IC_{50} values, which do not represent an absolute dissociation constant since they are dependent on the reference peptide. However, they allow a better comparison of binding affinities of different peptides than the results obtained by reconstitution or stabilization assays. Competition assays using purified MHC molecules employ soluble MHC molecules, which might have a slightly different structure compared to native molecules embedded in the cell membrane. Whole cell assays overcome these problems by using MHC molecules on the surface of intact cells. As an additional feature, mild acid treatment of the cells can be used to remove peptides from the MHC-binding groove. Cells are then incubated with a fluorescence-labeled reference peptide together with different concentrations of the peptides of interest (van der Burg et al., 1995). The amount of bound reference peptide can be measured by FACS analysis. As peptide-depleted MHC molecules have a short half-life and β_2-microglobulin dissociates rapidly, only membrane-anchored MHC-α-chains are left and, thus, this competition assay might also represent a reconstitution assay.

In classical reconstitution assays, cells expressing the appropriate HLA alleles are stripped by incubation at pH 3.2 for a short time, then the peptide of interest and a conformation-dependent monoclonal antibody are added. The difference in fluorescence intensity determined by FACS analysis between cells incubated with and without peptide after staining with fluorescence-labeled secondary antibody is used to determine peptide binding (Storkus et al., 1993).

The stabilization assay was first described by Stuber et al. (1992). Cell lines with defective peptide loading mechanisms, like the TAP-deficient murine cell line RMA-S (Ljunggren and Karre, 1985) or the human cell lines T2 (Cerundolo et al., 1990) or ST-EMO (de la Salle et al., 1994), display a low level of mostly empty MHC class I molecules on their surface due to the rapid turnover of empty MHC molecules. The addition of exogenous binding peptide to the cells, however, leads to stabilization and accumulation of MHC molecules, which can be determined using monoclonal antibodies and FACS analysis. Due to its simplicity, this assay can serve for screening large numbers of peptides. Moreover,

transfection of any human cell with cDNA of the immediate early protein ICP47 of the Herpes simplex virus generates a TAP-deficient phenotype and, thus, permits the extension of the stabilization assay to virtually any HLA allele (Gatfield et al., 1998).

3.2. Analysis of presentation

Since the first identification of a natural MHC class I ligand by tandem mass spectrometry (MSMS), this technique has become the bioanalytical method of choice for high sensitivity sequencing of natural MHC ligands. This has lead to the development of approaches that combine epitope prediction and HPLC-MS and MSMS for specific detection and identification of disease-associated MHC class I ligands. Brockman et al. (1999) used a simple main anchor-based epitope prediction to identify masses of candidate H2-K^b-ligands from *Trypanosoma cruzi*. The corresponding mass chromatograms showed in two cases significant peaks and subsequent HPLC-MSMS analysis revealed indeed two naturally processed peptides. A variant of this approach is the "Predict, Calibrate, Detect" (PCD) method (Schirle et al., 2000), which permits a more rapid and sensitive screening of tumor samples by calibrating the HPLC-MS system with synthetic analogs of predicted potential HLA-A2 ligands from tumor-associated proteins. HLA-A2 ligands from TAAs p53, CEA and MAGE-A1 have been identified using PCD (Schirle et al., 2000; Pascolo et al., 2001); the MAGE-A1-derived peptide has been verified as a CTL epitope using an HLA-A2 transgenic mouse model (Pascolo et al., 2001).

These approaches are not a substitute for testing peptide-specific recognition of tumor cells by CTL; however, they allow further focussing of these efforts to those peptides that are not only binding to the MHC molecule, but are actually processed and presented.

3.3. Analysis of T cell recognition

The final experimental verification of a predicted epitope is achieved by the identification of T cells, which specifically recognize the naturally processed epitope in an HLA-restricted fashion. One major source for T cells are peripheral blood mononuclear cells (PBMC) from patients or healthy donors, which are either already primed in vivo or which can be primed in vitro, respectively. In addition, immunization of various HLA class I and class II transgenic mice (reviewed in Faulkner et al., 1998 and Sonderstrup et al., 1999) can be used to generate specific T cells. T cell responses after stimulation can be measured by means of lytic activity, proliferation and cytokine production. Lytic activity of cytotoxic T lymphocytes has been analyzed since 1968 using the ^{51}Cr-release assay (Brunner et al., 1968; Cerottini and Brunner, 1974); however, this assay requires the cumbersome in vitro expansion of specific T cells from PBMCs. Nevertheless, it is still widely used. As an alternative, T cell proliferation has been investigated by radioactive ^{3}H-thymidine incorporation. To determine peptide-specific T cell responses, a sensitive proliferation assay was established in 1996. This assay avoids radioactivity and uses BrdU incorporation instead (Mayer et al., 1996). In our experience, this proliferation assay has a much higher sensitivity than the ^{51}Cr-release assay. The first quantitative assay for IL-2 production by activated T cells described in 1978 is still used today (Gillis et al., 1978). This sensitive method uses an IL-2-dependent mouse T cell line, CTLL, as the target cell population, which proliferates and incorporates ^{3}H-thymidine in response to IL-2 in culture medium. After IL-2- and IFNγ-specific antibodies became available, enzyme-linked immunosorbent assays (ELISA) were established to measure T cell responses (Gehman and Robb, 1984; Van der Meide et al., 1985; Troye-Blomberg et al., 1985). The advantages of ELISA procedures are their simple setup and the capacity to handle many samples in parallel.

None of the methods discussed so far is able to detect cytokine production by T cells at a single cell level. The first application to address this question was the ELISPOT assay reported in 1988 for enumerating IFNγ-secreting lymphocytes (Czerkinsky et al., 1988; Versteegen et al., 1988), originally developed for the detection of single specific antibody-secreting cells (Czerkinsky et al., 1983). In this assay, activated T cells are plated in anti-IFNγ antibody-coated wells; single IFNγ-secreting cells are detected as spots after staining trapped IFNγ with an insoluble dye generated by an enzyme coupled to the

second anti-IFNγ antibody. Peptide-specific T cells can be assessed in this way with higher sensitivity compared to the ^{51}Cr-release assay (Miyahira et al., 1995), even without the need for prior in vitro expansion, which may bias the result. ELISPOT has been used to measure TNFα secretion (Herr et al., 1996) or IFNγ secretion (Scheibenbogen et al., 1997; Schmittel et al., 1997) as markers for T cell activation. The tedious visual counting of spots was no longer necessary after the introduction of a computer-assisted image analysis system, which has made data analysis convenient, objective and suitable for handling large sample pools (Cui and Chang, 1997).

Another way to assess cytokine production at the single cell level is achieved by intracellular cytokine staining and flow cytometry analysis. After stimulation in the presence of monensin, T cells are fixed by paraformaldehyde and permeabilized by saponin to allow antibodies to penetrate through the cell membrane, the cytosol and the membranes of the endoplasmic reticulum and Golgi organelle. Monensin interrupts intracellular transport processes leading to an accumulation of the cytokine in the Golgi complex and, thus, an increase in the signal-to-noise ratio leads to a higher sensitivity (Sander et al., 1989,1991; Jung et al., 1993; Mascher et al., 1999; Pala et al., 2000). Intracellular cytokine staining allows high throughput of samples and multiparameter characterization of cytokine production as well as parallel detection of cell surface markers on a single cell basis without the need for prolonged in vitro culture and cloning. However, the sensitivity is limited by the fluorescence-activated cell-sorting analysis.

Similar to the intracellular cytokine staining approach is the cell surface affinity matrix technology (Manz et al., 1995). Instead of detecting cytokines in the cytoplasm, secreted cytokines are measured on live single cells. After antigenic restimulation, secreted IFNγ or other cytokines are captured by a cell surface affinity matrix (Brosterhus et al., 1999). This matrix comprises bispecific antibody–antibody conjugates directed against CD45 for cell membrane anchoring and against the cytokine of interest. Detection of secreted, matrix bound cytokine is achieved by fluorescence-labeled antibody and flow cytometry. In contrast to intracellular cytokine staining, this technology allows enrichment and isolation of living cells, which can be further cultured and analyzed.

Visualization of antigen-specific T cells is possible by staining T cell receptors with MHC/peptide tetrameric complexes (Altman et al., 1996; Gallimore et al., 1998; Murali-Krishna et al., 1998; Letvin et al., 1999). These tetramers are made by in vitro folding MHC heavy chain in the presence of β_2-microglobulin and a specific peptide ligand. The heavy chain is engineered in a way that the purified MHC-peptide complex can be biotinylated and then tetramerized via fluorochrome dye-coupled avidin. Tetramerization enhances the affinity of the αβTCRs to soluble peptide-loaded MHC molecules. By this method, a quantitation of T cells displaying TCRs specific for the analyzed peptide is possible; however, nothing can be said about the functionality of these T cells. The tetramer approach is well established for MHC class I molecules; however, detection of antigen-specific CD4$^+$ T cells with soluble MHC class II tetramers is now in progress (Crawford et al., 1998; Novak et al., 1999; Kotzin et al., 2000; Kuroda et al., 2000; Reichstetter et al., 2000; Kwok et al., 2000; Meyer et al., 2000). Application of MHC class I tetramers in the field of virus immunology has been reviewed by Doherty and Christensen (2000).

Probably, the most sensitive method for detecting T cell responses is the quantitation of induced cytokine mRNA after activation by quantitative real time reverse transcriptase-polymerase chain reaction (RT-PCR). This quantitative method is based on the 5′ nuclease assay in which the 5′ → 3′ endonuclease activity of the Taq DNA polymerase cleaves a hybridized sequence-specific fluorogenic internal oligonucleotide probe during PCR. Relative increases in fluorescence emission from the reporter dye are detected online in an analytical thermal cycler (Heid et al., 1996). This assay has been used to quantify cytokine mRNA expression (IL-2, IL-4, IL-10, TNFα, IFNγ) in previously frozen whole blood samples (Kruse et al., 1997). More recently, this very sensitive and highly reproducible method was adapted by normalizing copy numbers of cytokine mRNAs to CD8 mRNA copy numbers in order to analyze specific T cell responses in fresh PBMCs without any further stimulation (Kammula et al., 1999,2000). This way, it was possible to monitor the specific T cell response after peptide vaccination in melanoma patients. Though this is no single cell assay, it is very

sensitive, highly reproducible and permits a high sample throughput.

4. Outlook

While computational predictions have already become a familiar tool for many immunologists searching for T cell epitopes from immunologically relevant proteins, there is the constant need to use new data to update and extend the available programs. Reliable epitope prediction is still only available for a limited number of organisms and alleles because little or no information is available about the corresponding peptide specificities. However, several studies have shown both for MHC class I and II that the knowledge of which amino acids occupy those positions of the MHC polypeptide chain responsible for the interaction with peptide ligands make it possible, in some cases, to predict the peptide specificities for MHC molecules with as yet unknown motifs (Seeger et al., 1999; Sturniolo et al., 1999).

Furthermore, predictions for other characteristics of antigen processing, such as proteasomal cleavage specificities, are still in their infancy, with programs that are based on quantified cleavage data from immuno- and constitutive proteasomes not yet being publicly available. On the other hand, the development of prediction programs for MHC class II that are able to take antigen processing into consideration are still in the distant future.

Since the increasing usage of screening techniques, such as DNA microarrays (Lockhart et al., 1996) and serological analysis of recombinant cDNA expression libraries (SEREX) (Türeci et al., 1996), provide a rapidly growing pool of candidate protein antigens, interactive combinations of computational epitope predictions with these approaches, such as the recently described combination of DNA microarray analysis and TEPITOPE (Sturniolo et al., 1999), are expected to provide a rapidly growing number of epitopes for use in immunotherapy of a wide range of diseases.

However, while the described strategies combining computational prediction and experimental methods are likely to provide a rapidly increasing number of T cell epitopes as potential tools for therapeutic and diagnostic purposes, the identified epitopes still have to pass the ultimate test: they have to prove to be useful in the in vivo situation.

Acknowledgements

We would like to thank Hans-Georg Rammensee for continuous support and critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 510, project C7), the European Union QLG2-CT-1999-0713; BMH4-CT98-3589) and the Deutsche Krebshilfe (10-1258-St I).

References

Adams, H.P., Koziol, J.A., 1995. Prediction of binding to MHC class I molecules. J. Immunol. Methods 185, 181.

Altman, J.D., Moss, P.A., Goulder, P.J., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J., Davis, M.M., 1996. Phenotypic analysis of antigen-specific T lymphocytes. Science 274, 94.

Altuvia, Y., Sette, A., Sidney, J., Southwood, S., Margalit, H., 1997. A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. Hum. Immunol. 58, 1.

Brockman, A.H., Orlando, R., Tarleton, R.L., 1999. A new liquid chromatography/tandem mass spectrometric approach for the identification of class I major histocompatibility complex associated peptides that eliminates the need for bioassays. Rapid Commun. Mass Spectrom. 13, 1024.

Brosterhus, H., Brings, S., Leyendeckers, H., Manz, R.A., Miltenyi, S., Radbruch, A., Assenmacher, M., Schmitz, J., 1999. Enrichment and detection of live antigen-specific $CD4^+$ and $CD8^+$ T cells based on cytokine secretion. Eur. J. Immunol. 29, 4053.

Brunner, K.T., Mauel, J., Cerottini, J.C., Chapuis, B., 1968. Quantitative assay of the lytic action of immune lymphoid cells on 51-Cr-labelled allogeneic target cells in vitro; inhibition by isoantibody and by drugs. Immunology 14, 181.

Brusic, V., Rudy, G., Harrison, L.C., 1994. Prediction of MHC binding peptides using artificial networks. In: Sonier, R., Yu, X.H. (Eds.), Complex Systems: Mechanisms of Adaptation. IOS Press, Amsterdam, p. 253.

Brusic, V., Rudy, G., Honeyman, G., Hammer, J., Harrison, L., 1998. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. Bioinformatics 14, 121.

Celis, E., Tsai, V., Crimi, C., DeMars, R., Wentworth, P.A., Chesnut, R.W., Grey, H.M., Sette, A., Serra, H.M., 1994. Induction of anti-tumor cytotoxic T lymphocytes in normal humans using primary cultures and synthetic peptide epitopes. Proc. Natl. Acad. Sci. U. S. A. 91, 2105.

Cerottini, J.C., Brunner, K.T., 1974. Cell-mediated cytotoxicity, allograft rejection, and tumor immunity. Adv. Immunol. 18, 67.

Cerundolo, V., Alexander, J., Anderson, K., Lamb, C., Cresswell, P., McMichael, A., Gotch, F., Townsend, A., 1990. Presentation of viral antigen controlled by a gene in the major histocompatibility complex. Nature 345, 449.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A., Strominger, J.L., 1992. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature 358, 764.

Corr, M., Boyd, L.F., Padlan, E.A., Margulies, D.H., 1993. H-2Dd exploits a four residue peptide binding motif. J. Exp. Med. 178, 1877.

Craiu, A., Akopian, T., Goldberg, A., Rock, K.L., 1997. Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. Proc. Natl. Acad. Sci. U. S. A. 94, 10850.

Crawford, F., Kozono, H., White, J., Marrack, P., Kappler, J., 1998. Detection of antigen-specific T cells with multivalent soluble class II MHC covalent peptide complexes. Immunity 8, 675.

Cui, Y., Chang, L.J., 1997. Computer-assisted, quantitative cytokine enzyme-linked immunospot analysis of human immune effector cell function. BioTechniques 22, 1146.

Czerkinsky, C.C., Nilsson, L.A., Nygren, H., Ouchterlony, O., Tarkowski, A., 1983. A solid-phase enzyme-linked immunospot (ELISPOT) assay for enumeration of specific antibody-secreting cells. J. Immunol. Methods 65, 109.

Czerkinsky, C., Andersson, G., Ekre, H.P., Nilsson, L.A., Klareskog, L., Ouchterlony, O., 1988. Reverse ELISPOT assay for clonal analysis of cytokine production. I. Enumeration of gamma-interferon-secreting cells. J. Immunol. Methods 110, 29.

D'Amaro, J., Houbiers, J.G., Drijfhout, J.W., Brandt, R.M., Schipper, R., Bavinck, J.N., Melief, C.J., Kast, W.M., 1995. A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide-binding motifs. Hum. Immunol. 43, 13.

Daniel, S., Brusic, V., Caillat-Zucman, S., Petrovsky, N., Harrison, L., Riganelli, D., Sinigaglia, F., Gallazzi, F., Hammer, J., van Endert, P.M., 1998. Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. J. Immunol. 161, 617.

Davenport, M.P., Ho Shon, I., Hill, A.V., 1995. An empirical method for the prediction of T-cell epitopes. Immunogenetics 42, 392.

Davenport, M.P., Smith, K.J., Barouch, D., Reid, S.W., Bodnar, W.M., Willis, A.C., Hunt, D.F., Hill, A.V., 1997. HLA class I binding motifs derived from random peptide libraries differ at the COOH terminus from those of eluted peptides. J. Exp. Med. 185, 367.

De Groot, A.S., Jesdale, B.M., Szu, E., Schafer, J.R., Chicz, R.M., Deocampo, G., 1997. An interactive Web site providing major histocompatibility ligand predictions: application to HIV research. AIDS Res. Hum. Retroviruses 13, 529.

de la Salle, S.H., Hanau, D., Fricker, D., Urlacher, A., Kelly, A., Salamero, J., Powis, S.H., Donato, L., Bausinger, H., Laforet, M., Jeras, M., Spehner, D., Bieber, T., Falkenrodt, A., Cazenave, J.-P., Trowsdale, J., Tongio, M.M., 1994. Homozygous human TAP peptide transporter mutation in HLA class I deficiency. Science 265, 237.

DeLisi, C., Berzofsky, J.A., 1985. T-cell antigenic sites tend to be amphipathic structures. Proc. Natl. Acad. Sci. U. S. A. 82, 7048.

DiBrino, M., Parker, K.C., Shiloach, J., Knierman, M., Lukszo, J., Turner, R.V., Biddison, W.E., Coligan, J.E., 1993. Endogenous peptides bound to HLA-A3 possess a specific combination of anchor residues that permit identification of potential antigenic peptides. Proc. Natl. Acad. Sci. U. S. A. 90, 1508.

Doherty, P.C., Christensen, J.P., 2000. Accessing complexity: the dynamics of virus-specific T cell responses. Annu. Rev. Immunol. 18, 561.

Doolan, D.L., Southwood, S., Chesnut, R., Appella, E., Gomez, E., Richards, A., Higashimoto, Y.I., Maewal, A., Sidney, J., Gramzinski, R.A., Mason, C., Koech, D., Hoffman, S.L., Sette, A., 2000. HLA-DR-promiscuous T cell epitopes from *Plasmodium falciparum* pre-erythrocytic-stage antigens restricted by multiple HLA class II alleles. J. Immunol. 165, 1123.

Falk, K., Rötzschke, O., Stevanovic, S., Jung, G., Rammensee, H.G., 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351, 290.

Falk, K., Rötzschke, O., Stevanovic, S., Jung, G., Rammensee, H.G., 1994. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. Immunogenetics 39, 230.

Faulkner, L., Borysiewicz, L.K., Man, S., 1998. The use of human leucocyte antigen class I transgenic mice to investigate human immune function. J. Immunol. Methods 221, 1.

Fleckenstein, B., Kalbacher, H., Müller, C.P., Stoll, D., Halder, T., Jung, G., Wiesmüller, K.H., 1996. New ligands binding to the human leukocyte antigen class II molecule DRB1*0101 based on the activity pattern of an undecapeptide library. Eur. J. Biochem. 240, 71.

Gallimore, A., Glithero, A., Godkin, A., Tissot, A.C., Pluckthun, A., Elliott, T., Hengartner, H., Zinkernagel, R., 1998. Induction and exhaustion of lymphocytic choriomeningitis virus-specific cytotoxic T lymphocytes visualized using soluble tetrameric major histocompatibility complex class I-peptide complexes. J. Exp. Med. 187, 1383.

Gatfield, J., Lammert, E., Nickolaus, P., Münz, C., Rothenfusser, S., Fisch, P., Stevanovic, S., Schild, H., Rammensee, H.G., Arnold, D., 1998. Cell lines transfected with the TAP inhibitor ICP47 allow testing peptide binding to a variety of HLA class I molecules. Int. Immunol. 10, 1665.

Gehman, L.O., Robb, R.J., 1984. An ELISA-based assay for quantitation of human interleukin 2. J. Immunol. Methods 74, 39.

Gillis, S., Ferm, M.M., Ou, W., Smith, K.A., 1978. T cell growth factor: parameters of production and a quantitative microassay for activity. J. Immunol. 120, 2027.

Godkin, A.J., Davenport, M.P., Willis, A., Jewell, D.P., Hill, A.V., 1998. Use of complete eluted peptide sequence data from HLA-DR and -DQ molecules to predict T cell epitopes,

and the influence of the nonbinding terminal regions of ligands in epitope selection. J. Immunol. 161, 850.

Gulukota, K., Sidney, J., Sette, A., DeLisi, C., 1997. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. J. Mol. Biol. 267, 1258.

Halder, T., Pawelec, G., Kirkin, A.F., Zeuthen, J., Meyer, H.E., Kun, L., Kalbacher, H., 1997. Isolation of novel HLA-DR restricted potential tumor-associated antigens from the melanoma cell line FM3. Cancer Res. 57, 3238.

Hammer, J., Belunis, C., Bolin, D., Papadopoulos, J., Walsky, R., Higelin, J., Danho, W., Sinigaglia, F., Nagy, Z.A., 1994a. High-affinity binding of short peptides to major histocompatibility complex class II molecules by anchor combinations. Proc. Natl. Acad. Sci. U. S. A. 91, 4456.

Hammer, J., Bono, E., Gallazzi, F., Belunis, C., Nagy, Z., Sinigaglia, F., 1994b. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. J. Exp. Med. 180, 2353.

Hammer, J., Gallazzi, F., Bono, E., Karr, R.W., Guenot, J., Valsasnini, P., Nagy, Z.A., Sinigaglia, F., 1995. Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid arthritis association. J. Exp. Med. 181, 1847.

Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M., 1996. Real time quantitative PCR. Genome Res. 6, 986.

Herr, W., Schneider, J., Lohse, A.W., Meyer zum Buschenfelde, K.H., Wölfel, T., 1996. Detection and quantification of blood-derived $CD8^+$ T lymphocytes secreting tumor necrosis factor alpha in response to HLA-A2.1-binding melanoma and viral peptide antigens. J. Immunol. Methods 191, 131.

Holzhütter, H.G., Frommel, C., Kloetzel, P.M., 1999. A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. J. Mol. Biol. 286, 1251.

Honeyman, M.C., Brusic, V., Stone, N.L., Harrison, L.C., 1998. Neural network-based prediction of candidate T-cell epitopes. Nat. Biotechnol. 16, 966.

Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., Engelhard, V.H., 1992a. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 255, 1261.

Hunt, D.F., Michel, H., Dickinson, T.A., Shabanowitz, J., Cox, A.L., Sakaguchi, K., Appella, E., Grey, H.M., Sette, A., 1992b. Peptides presented to the immune system by the murine class II major histocompatibility complex molecule $I\text{-}A^d$. Science 256, 1817.

Jensen, P.E., 1991. Enhanced binding of peptide antigen to purified class II major histocompatibility glycoproteins at acidic pH. J. Exp. Med. 174, 1111.

Jung, T., Schauer, U., Heusser, C., Neumann, C., Rieger, C., 1993. Detection of intracellular cytokines by flow cytometry. J. Immunol. Methods 159, 197.

Kammula, U.S., Lee, K.H., Riker, A.I., Wang, E., Ohnmacht, G.A., Rosenberg, S.A., Marincola, F.M., 1999. Functional analysis of antigen-specific T lymphocytes by serial measurement of gene expression in peripheral blood mononuclear cells and tumor specimens. J. Immunol. 163, 6867.

Kammula, U.S., Marincola, F.M., Rosenberg, S.A., 2000. Real-time quantitative polymerase chain reaction assessment of immune reactivity in melanoma patients after tumor peptide vaccination. J. Natl. Cancer Inst. 92, 1336.

Kessler, J.H., Beekman, N.J., Bres-Vloemans, S.A., Verdijk, P., van Veelen, P.A., Kloosterman-Joosten, A.M., Vissers, D.C., ten Bosch, G.J., Kester, M.G., Sijts, A., Wouter, D.J., Ossendorp, F., Offringa, R., Melief, C.J., 2001. Efficient identification of novel HLA-A(*)0201-presented cytotoxic T lymphocyte epitopes in the widely expressed tumor antigen PRAME by proteasome-mediated digestion analysis. J. Exp. Med. 193, 73.

Kotzin, B.L., Falta, M.T., Crawford, F., Rosloniec, E.F., Bill, J., Marrack, P., Kappler, J., 2000. Use of soluble peptide-DR4 tetramers to detect synovial T cells specific for cartilage antigens in patients with rheumatoid arthritis. Proc. Natl. Acad. Sci. U. S. A. 97, 291.

Kruse, N., Pette, M., Toyka, K., Rieckmann, P., 1997. Quantification of cytokine mRNA expression by RT PCR in samples of previously frozen blood. J. Immunol. Methods 210, 195.

Kubo, R.T., Sette, A., Grey, H.M., Appella, E., Sakaguchi, K., Zhu, N.Z., Arnott, D., Sherman, N., Shabanowitz, J., Michel, H., 1994. Definition of specific peptide motifs for four major HLA-A alleles. J. Immunol. 152, 3913.

Kuroda, M.J., Schmitz, J.E., Lekutis, C., Nickerson, C.E., Lifton, M.A., Franchini, G., Harouse, J.M., Cheng-Mayer, C., Letvin, N.L., 2000. Human immunodeficiency virus type 1 envelope epitope-specific CD4(+) T lymphocytes in simian/human immunodeficiency virus-infected and vaccinated rhesus monkeys detected using a peptide-major histocompatibility complex class II tetramer. J. Virol. 74, 8751.

Kwok, W.W., Liu, A.W., Novak, E.J., Gebe, J.A., Ettinger, R.A., Nepom, G.T., Reymond, S.N., Koelle, D.M., 2000. HLA-DQ tetramers identify epitope-specific T cells in peripheral blood of herpes simplex virus type 2-infected individuals: direct detection of immunodominant antigen-responsive cells. J. Immunol. 164, 4244.

Leighton, J., Sette, A., Sidney, J., Appella, E., Ehrhardt, C., Fuchs, S., Adorini, L., 1991. Comparison of structural requirements for interaction of the same peptide with I-Ek and I-Ed molecules in the activation of MHC class II-restricted T cells. J. Immunol. 147, 198.

Letvin, N.L., Schmitz, J.E., Jordan, H.L., Seth, A., Hirsch, V.M., Reimann, K.A., Kuroda, M.J., 1999. Cytotoxic T lymphocytes specific for the simian immunodeficiency virus. Immunol. Rev. 170, 127.

Ljunggren, H.G., Karre, K., 1985. Host resistance directed selectively against H-2-deficient lymphoma variants. Analysis of the mechanism. J. Exp. Med. 162, 1745.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E.L., 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14, 1675.

Lu, J., Celis, E., 2000. Use of two predictive algorithms of the world wide web for the identification of tumor-reactive T-cell epitopes. Cancer Res. 60, 5223.

Mallios, R.R., 1999. Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-algorithm. Bioinformatics 15, 432.

Manici, S., Sturniolo, T., Imro, M.A., Hammer, J., Sinigaglia, F., Noppen, C., Spagnoli, G., Mazzi, B., Bellone, M., Dellabona, P., Protti, M.P., 1999. Melanoma cells present a MAGE-3 epitope to CD4(+) cytotoxic T cells in association with histocompatibility leukocyte antigen DR11. J. Exp. Med. 189, 871.

Manz, R., Assenmacher, M., Pfluger, E., Miltenyi, S., Radbruch, A., 1995. Analysis and sorting of live cells according to secreted molecules, relocated to a cell-surface affinity matrix. Proc. Natl. Acad. Sci. U. S. A. 92, 1921.

Marshall, K.W., Wilson, K.J., Liang, J., Woods, A., Zaller, D., Rothbard, J.B., 1995. Prediction of peptide affinity to HLA DRB1*0401. J. Immunol. 154, 5927.

Mascher, B., Schlenke, P., Seyfarth, M., 1999. Expression and kinetics of cytokines determined by intracellular staining using flow cytometry. J. Immunol. Methods 223, 115.

Mayer, S., Scheibenbogen, C., Lee, K.H., Keilholz, W., Stevanovic, S., Rammensee, H.G., Keilholz, U., 1996. A sensitive proliferation assay to determine the specific T cell response against HLA-A2.1-binding peptides. J. Immunol. Methods 197, 131.

Meyer, A.L., Trollmo, C., Crawford, F., Marrack, P., Steere, A.C., Huber, B.T., Kappler, J., Hafler, D.A., 2000. Direct enumeration of Borrelia-reactive CD4 T cells ex vivo by using MHC class II tetramers. Proc. Natl. Acad. Sci. U. S. A. 97, 11433.

Milik, M., Sauer, D., Brunmark, A.P., Yuan, L., Vitiello, A., Jackson, M.R., Peterson, P.A., Skolnick, J., Glass, C.A., 1998. Application of an artificial neural network to predict specific class I MHC binding peptide sequences. Nat. Biotechnol. 16, 753.

Miyahira, Y., Murata, K., Rodriguez, D., Rodriguez, J.R., Esteban, M., Rodrigues, M.M., Zavala, F., 1995. Quantification of antigen specific CD8+T cells using an ELISPOT assay. J. Immunol. Methods 181, 45.

Momburg, F., Roelse, J., Howard, J.C., Butcher, G.W., Hammerling, G.J., Neefjes, J.J., 1994. Selectivity of MHC-encoded peptide transporters from human, mouse and rat. Nature 367, 648.

Murali-Krishna, K., Altman, J.D., Suresh, M., Sourdive, D.J., Zajac, A.J., Miller, J.D., Slansky, J., Ahmed, R., 1998. Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. Immunity 8, 177.

Novak, E.J., Liu, A.W., Nepom, G.T., Kwok, W.W., 1999. MHC class II tetramers identify peptide-specific human CD4(+) T cells proliferating in response to influenza A antigen. J. Clin. Invest. 104, R63.

Nussbaum, A.K., Dick, T.P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D.H., Huber, R., Rammensee, H.G., Schild, H., 1998. Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. Proc. Natl. Acad. Sci. U. S. A. 95, 12504.

Nussbaum, A.K., Kuttler, C., Hadeler, K.P., Rammensee, H.G., Schild, H., 2001. PapRoC: a predicition algorithm for proteasomal cleavages available on the WWW. Immunogenetics 53, 87.

Olsen, A.C., Pedersen, L.O., Hansen, A.S., Nissen, M.H., Olsen, M., Hansen, P.R., Holm, A., Buus, S., 1994. A quantitative assay to measure the interaction between immunogenic peptides and purified class I major histocompatibility complex molecules. Eur. J. Immunol. 24, 385.

Pala, P., Hussell, T., Openshaw, P.J., 2000. Flow cytometric measurement of intracellular cytokines. J. Immunol. Methods 243, 107.

Parner, E.G., Harty, J.T., Bevan, M.J., 1991. Precise prediction of a dominant class I MHC-restricted epitope of *Listeria monocytogenes*. Nature 353, 852.

Parker, K.C., Bednarek, M.A., Coligan, J.E., 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152, 163.

Pascolo, S., Schirle, M., Gückel, B., Dumrese, T., Stumm, S., Kayser, S., Moris, A., Wallwiener, D., Rammensee, H.G., Stevanovic, S., 2001. A MAGE-A1 HLA-A*0201 epitope identified by mass spectrometry. Cancer Res. 61, 4072.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., Stevanovic, S., 1999. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50, 213.

Rongcun, Y., Salazar-Onfray, F., Charo, J., Malmberg, K.J., Evrin, K., Maes, H., Kono, K., Hising, C., Petersson, M., Larsson, O., Lan, L., Appella, E., Sette, A., Celis, E., Kiessling, R., 1999. Identification of new HER2/neu-derived peptide epitopes that can elicit specific CTL against autologous and allogeneic carcinomas and melanomas. J. Immunol. 163, 1037.

Reichstetter, S., Ettinger, R.A., Liu, A.W., Gebe, J.A., Nepom, G.T., Kwok, W.W., 2000. Distinct T cell interactions with HLA class II tetramers characterize a spectrum of TCR affinities in the human antigen-specific T cell response. J. Immunol. 165, 6994.

Rothbard, J.B., Taylor, W.R., 1988. A sequence pattern common to T cell epitopes. EMBO J. 7, 93.

Rötzschke, O., Falk, K., Stevanovic, S., Jung, G., Walden, P., Rammensee, H.G., 1991. Exact prediction of a natural T cell epitope. Eur. J. Immunol. 21, 2891.

Rudensky, A.Y., Preston-Hurlburt, P., al-Ramadi, B.K., Rothbard, J., Janeway, C.A., 1992. Truncation variants of peptides isolated from MHC class II molecules suggest sequence motifs. Nature 359, 429.

Ruppert, J., Sidney, J., Celis, E., Kubo, R.T., Grey, H.M., Sette, A., 1993. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. Cell 74, 929.

Sander, B., Cardell, S., Heremans, H., Andersson, U., Moller, G., 1989. Detection of individual interleukin 4- and gamma interferon-producing murine spleen cells after activation with T-cell mitogens. Scand. J. Immunol. 30, 315.

Sander, B., Andersson, J., Andersson, U., 1991. Assessment of cytokines by immunofluorescence and the paraformaldehyde-saponin procedure. Immunol. Rev. 119, 65.

Scheibenbogen, C., Lee, K.H., Mayer, S., Stevanovic, S., Moe-

bius, U., Herr, W., Rammensee, H.G., Keilholz, U., 1997. A sensitive ELISPOT assay for detection of CD8+T lymphocytes specific for HLA class I-binding peptide epitopes derived from influenza proteins in the blood of healthy donors and melanoma patients. Clin. Cancer Res. 3, 221.

Schirle, M., Keilholz, W., Weber, B., Gouttefangeas, C., Dumrese, T., Becker, H.D., Stevanovic, S., Rammensee, H.G., 2000. Identification of tumor-associated MHC class I ligands by a novel T cell-independent approach. Eur J. Immunol. 30, 2216.

Schmittel, A., Keilholz, U., Scheibenbogen, C., 1997. Evaluation of the interferon-gamma ELISPOT-assay for quantification of peptide specific T lymphocytes from peripheral blood. J. Immunol. Methods 210, 167.

Schueler-Furman, O., Altuvia, Y., Sette, A., Margalit, H., 2000. Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. Protein Sci. 9, 1838.

Seeger, F.H., Schirle, M., Gatfield, J., Arnold, D., Keilholz, W., Nickolaus, P., Rammensee, H.G., Stevanovic, S., 1999. The HLA-A*6601 peptide motif: prediction by pocket structure and verification by peptide analysis. Immunogenetics 49, 571.

Sette, A., Buus, S., Appella, E., Smith, J.A., Chesnut, R., Miles, C., Colon, S.M., Grey, H.M., 1989. Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. Proc. Natl. Acad. Sci. U. S. A. 86, 3296.

Sette, A., Sidney, J., del Guercio, M.F., Southwood, S., Ruppert, J., Dahlberg, C., Grey, H.M., Kubo, R.T., 1994. Peptide binding to the most frequent HLA-A class I alleles measured by quantitative molecular binding assays. Mol. Immunol. 31, 813.

Sonderstrup, G., Cope, A.P., Patel, S., Congia, M., Hain, N., Hall, F.C., Parry, S.L., Fugger, L.H., Michie, S., McDevitt, H.O., 1999. HLA class II transgenic mice: models of the human CD4+T-cell immune response. Immunol. Rev. 172, 335.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., Sette, A., 1998. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 160, 3363.

Stoltze, L., Dick, T.P., Deeg, M., Pömmerl, B., Rammensee, H.G., Schild, H., 1998. Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and -independent proteolytic activities. Eur. J. Immunol. 28, 4029.

Storkus, W.J., Zeh, H.J., Salter, R.D., Lotze, M.T., 1993. Identification of T-cell epitopes: rapid isolation of class I-presented peptides from viable cells by mild acid elution. J. Immunother. 14, 94.

Stryhn, A., Pedersen, L.O., Romme, T., Holm, C.B., Holm, A., Buus, S., 1996. Peptide binding specificity of major histocompatibility complex class I resolved into an array of apparently independent subspecificities: quantitation by peptide libraries and improved prediction of binding. Eur. J. Immunol. 26, 1911.

Stryhn, A., Pedersen, L.O., Buus, S., 1997. Biochemical analysis of peptide binding to MHC class I. In: Lefkovits, I. (Ed.), Immunology Methods Manual. Academic Press, London, p. 637.

Stuber, G., Modrow, S., Hoglund, P., Franksson, L., Elvin, J., Wolf, H., Karre, K., Klein, G., 1992. Assessment of major histocompatibility complex class I interaction with Epstein-Barr virus and human immunodeficiency virus peptides by elevation of membrane H-2 and HLA in peptide loading-deficient cells. Eur. J. Immunol. 22, 2697.

Stuber, G., Leder, G.H., Storkus, W.T., Lotze, M.T., Modrow, S., Szekely, L., Wolf, H., Klein, E., Karre, K., Klein, G., 1994. Identification of wild-type and mutant p53 peptides binding to HLA-A2 assessed by a peptide loading-deficient cell line assay and a novel major histocompatibility complex class I peptide binding assay. Eur. J. Immunol. 24, 765.

Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Türeci, Ö., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M.P., Sinigaglia, F., Hammer, J., 1999. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat. Biotechnol. 17, 555.

Sun, Y., Song, M., Stevanovic, S., Jankowiak, C., Paschen, A., Rammensee, H.G., Schadendorf, D., 2000. Identification of a new HLA-A(*)0201-restricted T-cell epitope from the tyrosinase-related protein 2 (TRP2) melanoma antigen. Int. J. Cancer 87, 399.

Toes, R.E., Nussbaum, A.K., Degermann, S., Schirle, M., Emmerich, N.P., Kraft, M., Laplace, C., Zwinderman, A., Dick, T.P., Müller, J., Schönfisch, B., Schmid, C., Fehling, H.J., Stevanovic, S., Rammensee, H.G., Schild, H., 2001. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J. Exp. Med. 194, 1.

Tompkins, S.M., Rota, P.A., Moore, J.C., Jensen, P.E., 1993. A europium fluoroimmunoassay for measuring binding of antigen to class II MHC glycoproteins. J. Immunol. Methods 163, 209.

Touloukian, C.E., Leitner, W.W., Topalian, S.L., Li, Y.F., Robbins, P.F., Rosenberg, S.A., Restifo, N.P., 2000. Identification of a MHC class II-restricted human gp100 epitope using DR4-IE transgenic mice. J. Immunol. 164, 3535.

Troye-Blomberg, M., Andersson, G., Stoczkowska, M., Shabo, R., Romero, P., Patarroyo, E., Wigzell, H., Perlmann, P., 1985. Production of IL 2 and IFN-gamma by T cells from malaria patients in response to *Plasmodium falciparum* or erythrocyte antigens in vitro. J. Immunol. 135, 3498.

Türeci, Ö., Sahin, U., Schobert, I., Koslowski, M., Schmitt, H., Schild, H.J., Stenner, F., Seitz, G., Rammensee, H.G., Pfreundschuh, M., 1996. The SSX-2 gene, which is involved in the t(X;18) translocation of synovial sarcomas, codes for the human tumor antigen HOM-MEL-40. Cancer Res. 56, 4766.

Udaka, K., Wiesmüller, K.H., Kienle, S., Jung, G., Tamamura, H., Yamagishi, H., Okumura, K., Walden, P., Suto, T., Kawasaki, T., 2000. An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries. Immunogenetics 51, 816.

van der Burg, S.H., Ras, E., Drijfhout, J.W., Benckhuijsen, W.E., Bremers, A.J., Melief, C.J., Kast, W.M., 1995. An HLA class I peptide-binding assay based on competition for binding to class I molecules on intact human B cells. Identification of

conserved HIV-1 polymerase peptides binding to HLA-A*0301. Hum. Immunol. 44, 189.

Van der Meide, P.H., Dubbeld, M., Schellekens, H., 1985. Monoclonal antibodies to human immune interferon and their use in a sensitive solid-phase ELISA. J. Immunol. Methods 79, 293.

Versteegen, J.M., Logtenberg, T., Ballieux, R.E., 1988. Enumeration of IFN-gamma-producing human lymphocytes by spot-ELISA. A method to detect lymphokine-producing lymphocytes at the single-cell level. J. Immunol. Methods 111, 25.

Vissers, J.L., De, V., Schreurs, I., Engelen, M.W., Oosterwijk, L.P., Figdor, E., Adema, C.G., 1999. The renal cell carcinoma-associated antigen G250 encodes a human leukocyte antigen (HLA)-A2.1-restricted epitope recognized by cytotoxic T lymphocytes. Cancer Res. 59, 5554.

Vogt, A.B., Kropshofer, H., Kalbacher, H., Kalbus, M., Rammensee, H.G., Coligan, J.E., Martin, R., 1994. Ligand motifs of HLA-DRB5*0101 and DRB1*1501 molecules delineated from self-peptides. J. Immunol. 153, 1665.

Vonderheide, R.H., Hahn, W.C., Schultze, J.L., Nadler, L.M., 1999. The telomerase catalytic subunit is a widely expressed tumor-associated antigen recognized by cytotoxic T lymphocytes. Immunity 10, 673.

Zarour, H.M., Kirkwood, J.M., Kierstead, L.S., Herr, W., Brusic, V., Slingluff, C.L., Sidney, J., Sette, A., Storkus, W.J., 2000a. Melan-A/MART-1(51–73) represents an immunogenic HLA-DR4-restricted epitope recognized by melanoma-reactive CD4(+) T cells. Proc. Natl. Acad. Sci. U. S. A. 97, 400.

Zarour, H.M., Storkus, W.J., Brusic, V., Williams, E., Kirkwood, J.M., 2000b. NY-ESO-1 encodes DRB1*0401-restricted epitopes recognized by melanoma-reactive CD4+T cells. Cancer Res. 60, 4946.

Zeh, H.J., Leder, G.H., Lotze, M.T., Salter, R.D., Tector, M., Stuber, G., Modrow, S., Storkus, W.J., 1994. Flow-cytometric determination of peptide-class I complex formation. Identification of p53 peptides that bind to HLA-A2. Hum. Immunol. 39, 79.

Generation of Human Cytotoxic T Cells Specific for Human Carcinoembryonic Antigen Epitopes From Patients Immunized With Recombinant Vaccinia-CEA Vaccine

*Kwong Y. Tsang, Sam Zaremba, Carol A. Nieroda, Ming Z. Zhu, J. Michael Hamilton, Jeffrey Schlom**

Background: **The human carcinoembryonic antigen (CEA), which is expressed in several cancer types, is a potential target for specific immunotherapy using recombinant vaccines. Previous studies have shown that when the CEA gene is placed into vaccinia virus, the recombinant vaccine (rV-CEA) can elicit T-cell responses in both rodents and nonhuman primates.** ***Purpose:*** **Our objective was to determine if rV-CEA could elicit CEA-specific T-cell responses in humans with appropriate human leukocyte antigen (HLA) motifs.** ***Methods:*** **Peripheral blood lymphocytes (PBLs) obtained from patients with metastatic carcinoma, both before and after vaccination with rV-CEA, were analyzed for T-cell response to specific 9- to 11-mer CEA peptides selected to conform to human HLA class I-A2 motifs.** ***Results:*** **While little or no T-cell growth was seen from preimmunization PBLs of patients pulsed with CEA peptides and interleukin 2 (IL-2), T-cell lines were obtained from PBLs of patients after vaccination with one to three cycles of stimulation. Cytolytic T-cell lines from three HLA-A2 patients were established with a 9-amino acid peptide (CAP-1), and the $CD8^+/CD4^+$ double-positive T-cell line (V24T) was chosen for detailed analysis. When autologous Epstein-Barr virus (EBV)-transformed B cells were either incubated with CAP-1 peptide or transduced with the CEA gene using a retroviral vector, they were lysed by the V24T cell line, but allogeneic non-A2 EBV-transformed B cells were not. The SW403 human colon carcinoma cell line, which is CEA positive and HLA-A2 positive, was also lysed by the V24T cell line, while two non–HLA-A2 CEA-positive colon carcinoma cell lines were not. To further confirm the class I HLA-A2 restricted nature of the V24T cytotoxicity, the non–HLA-A2 SW837 CEA-expressing colon carcinoma cell line was infected with a recombinant vaccinia virus expressing the HLA class I-A2 gene, and it became susceptible to V24T lysis. Cells infected with vector alone were not lysed.** ***Conclusions:*** **This study demonstrates for the first time (*a*) the ability to generate a human cytolytic T-cell response to specific epitopes of CEA, (*b*) the class I HLA-A2 restricted nature of the T-cell mediated lysis, and (*c*) the ability of human tumor cells to endogenously process CEA to present a specific CEA peptide in the context of major histocompatibility complex for T-cell–mediated lysis.** ***Implications:*** **These findings have implications in the development of specific second-generation cancer immunotherapy protocols. [J Natl Cancer Inst 87:982-990, 1995]**

The identification and selection of antigens and specific epitopes as targets for active immunotherapy approaches to human cancer are now in a dynamic phase. Specific peptides that bind human major histocompatibility complex (MHC) molecules have now been identified for melanoma-associated antigens (*1-4*). The identification of human carcinoma-associated antigens and epitopes that can be recognized by human T cells is also currently under active investigation. Molecules, such as prostate specific antigen (PSA) (*5,6*), c-erbB/2 (*7*), MUC-1 (*8*), point mutated ras (*9-11*), point mutated p53 (*12*), and carcinoembryonic antigen (CEA) (*13-15*) are among such candidates.

In humans, CEA is extensively expressed on the vast majority of colorectal, gastric, and pancreatic carcinomas as well as approximately 50% of breast cancers and 70% of non–small-cell lung cancers (*16*). CEA is also expressed, to some extent, on normal colon epithelium and in some fetal tissue (*16*). The CEA gene has been sequenced and shown to be part of the human immunoglobulin gene superfamily (*16,17*) and, thus, shares some homology with other molecules found on normal human tissues. At the amino acid level, CEA shares approximately 70% homology with nonspecific cross-reacting antigen (NCA), which is found on normal granulocytes (*16*).

The immunogenicity of CEA in humans is, at best, controversial. Several studies (*18,19*) claim antibodies to CEA in patients, while other investigators report these observations are artifacts (*20-22*). No reports of the presence or absence of human T-cell responses to CEA exist.

**Affiliations of authors:* K. Y. Tsang, S. Zaremba, C. A. Nieroda, M. Z. Zhu, J. Schlom (Laboratory of Tumor Immunology and Biology, Division of Cancer Biology, Diagnosis, and Centers), J. M. Hamilton (NCI-Navy Medical Oncology Branch, Division of Cancer Treatment), National Cancer Institute, Bethesda, Md.

Correspondence to: Jeffrey Schlom, Ph.D., National Institutes of Health, Bldg. 10, Rm. 8B07, Bethesda, MD 20892.

See "Notes" section following "References."

One strategy that is being pursued to determine if T-cell responses to CEA can be induced in carcinoma patients is to place the CEA gene into vaccinia virus. Vaccinia was chosen as a vector for several reasons. Among these are (*a*) its wide use in humans in the eradication of smallpox; (*b*) its ability to infect a wide range of cells, including professional antigen-presenting cells (APCs), and express the product of an inserted gene such that it has the potential to be processed in the context of class I and/or class II MHC molecules; and (*c*) that animal model studies have shown that the use of a recombinant human CEA vaccinia virus (designated rV-CEA) is superior to the use of soluble CEA in the induction of antitumor effects on established CEA-expressing tumors (*13*). These findings correlated with the appearance of CEA-specific cytotoxic T-lymphocytes (CTLs) in rV-CEA–inoculated animals (*13*). rV-CEA also has been administered to rhesus monkeys and has been shown to induce CEA-specific T-cell responses with no toxicity (*14*).

It is important to emphasize, however, that experimental model results should have extremely limited extrapolation to potential human immune T-cell responses. Human CEA is a foreign gene in both mice and nonhuman primates. However, the more important point to consider is whether human APCs, including tumor cells, will process CEA in such a manner as to present specific CEA peptides in the context of human MHC for human T-cell recognition. Since mouse and nonhuman primate MHC-binding motifs are different from human motifs, studies in animal models cannot answer the question of T-cell immunogenicity in humans. Even the use of CEA-transgenic mice could not answer these questions, because they would possess murine MHC motifs. Thus, while animal model studies were conducted to demonstrate that rV-CEA can infect mammalian cells in vivo to such a level as to induce immune responses and to demonstrate the lack of toxicity, only clinical trials can adequately answer the question of the potential ability of rV-CEA to induce CEA-specific human T-cell responses.

A phase I clinical trial approved by the National Cancer Institute (NCI) Institutional Review Board and conducted by the NCI-Navy Oncology Branch involving the use of rV-CEA in 26 patients with metastatic carcinoma (gastrointestinal, lung, and breast) has recently been completed (*23*). No toxicity was observed other than that usually seen with the smallpox vaccine. A maximum tolerated dose was not achieved, even in the group that received rV-CEA injections of 10^7 plaque forming units (pfu) once a month for 3 months. While T-cell responses to vaccinia virus were observed (*23*), no primary T-cell lymphoproliferative response was observed when soluble CEA protein was presented to peripheral blood lymphocytes (PBLs) obtained prior to or after rV-CEA vaccination.

In an effort to further analyze CEA-specific T-cell responses as a result of rV-CEA vaccination, peptides reflecting potential human class I T-cell epitopes were selected and used with interleukin 2 (IL-2) to stimulate PBLs of patients before and after vaccination with rV-CEA. What has emerged is the first evidence for the induction of CEA-specific CTL responses in patients after rV-CEA vaccination.

Materials and Methods

rV-CEA

A recombinant vaccinia virus expressing CEA (rV-CEA) was generated by Therion Biologics Corporation (Cambridge, MA) using the methods described (*24*). The CEA gene was isolated as a complementary DNA (cDNA) clone from a human colon carcinoma cell cDNA library. The CEA cDNA was inserted, under the control of the vaccinia 40K promoter (*24*), into the *Hind* III M region of the genome of the attenuated strain of vaccinia virus (Wyeth strain).

Cell Cultures

Colorectal carcinoma cell lines SW403 (human leukocyte antigen [HLA]-A2 and HLA-A3), HT-29 (HLA-A1 and HLA-A9), SW837 (HLA-A19, -), and SW1417 (HLA-A3, -) were purchased from American Type Culture Collection (Rockville, MD). The cultures were mycoplasma free and were maintained in complete medium (Dulbecco's modified Eagle medium; Life Technologies, Inc. [GIBCO BRL], Grand Island, NY) supplemented with 10% fetal bovine serum (FBS), 2 m*M* glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin (Life Technologies, Inc.). The T2 cell line (transport deletion mutant) (*25*) was provided by Dr. Peter Cresswell (Yale University School of Medicine, New Haven, CT) and was maintained in Iscove's modified Dulbecco's medium (IMDM) containing 10% FBS.

Epstein-Barr virus (EBV)-transformed B cell lines designated as B-Vac24 and B-Vac01, and the B-Vac24 transfected with a retroviral vector containing the CEA gene [designated as B-Vac24(CEA)] were maintained in RPMI-1640 medium supplemented with 10% pooled human AB serum (Pel Freeze Clinical System, Brown Deer, WI), 2 m*M* glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin (Life Technologies, Inc.).

Peptide Synthesis

The peptide sequence of CEA was scanned for matches to the consensus motifs for HLA-A2 and HLA-A3 binding peptides. HLA-A2 and HLA-A3 alleles were chosen, since they are the most commonly expressed class I alleles. Peptides (9-, 10- and 11-mers) were selected for synthesis if they (*a*) conformed to the respective consensus motifs and (*b*) diverged sufficiently from NCA and human biliary glycoprotein (BGP), so that a response to these antigens would not be anticipated. A peptide that corresponded to the CAP-1 peptide (*see* below) after optimal alignment with NCA and CEA was also synthesized and designated NCA-1. Syntheses were performed on a peptide synthesizer (model 432A; Applied Biosystems, Foster City, CA), and products were dissolved in aqueous solution, sterile filtered, and frozen at −70 °C at a concentration of 2 mg/mL. The purity of the peptides was greater than 90% as analyzed by high-performance liquid chromatography (HPLC). The CEA peptides are listed in Table 1.

Introduction of the CEA cDNA in EBV-Immortalized B-Cell Lines

Since tumors from rV-CEA vaccinated patients were not available, EBV-transformed B cells from these patients were used as autologous targets for T cells. B-cell lines were generated by a standard method (*26*) using B95-8 marmoset cell line supernatant containing EBV. Human pooled AB serum was used in all cell cultures in this study. EBV-immortalized B-cell lines were transduced with a retroviral expression construct of CEA (*27*). Transduction was performed by cocultivation of EBV-immortalized B cells with productively transduced amphotropic retroviral packaging cell line PA317-CEA as described by Tsang et al. (*28*). EBV-immortalized B-cell transductants were selected in medium containing G418 (geneticin) at an active concentration of 0.7 mg/mL. The purpose of the generation of immortalized cells was to have a continuous supply of targets to analyze cytotoxic T-cell responses.

Generation of T-Cell Lines

Peripheral blood mononuclear cells (PBMCs) were obtained from heparinized blood from patients with metastatic carcinoma who were enrolled in a phase I trial employing rV-CEA (*13,14*). All experiments involving patient materials were conducted according to NIH guidelines and written, informed consent was obtained from all patients. PBMCs were obtained prior to and after administration of injections of rV-CEA given once a month for 3 months at 2 ×

Table 1. Binding of CEA peptides to the HLA class I-A2 molecule

Peptide	Amino acid position in CEA	Sequence	Predicted binding to HLA class I-A2*	T2 binding assay†	
				Experiment 1	Experiment 2
CAP-1	571-579	YLSGANLNL	P	561	806
CAP-2	555-564	VLYGPDTPII	P	515	796
CAP-3	87-96	TLHVIKSDLV	P	480	ND
CAP-4	1-11	KLTIESTPFNV	P	441	ND
CAP-5	345-354	TLLSVTRNDV	P	405	ND
CAP-6	19-28	LLVHNLPQHL	P	381	ND
CAP-7	27-35	HLFGYSWYK	N	326	ND
CAP-8	523-532	TLFNVTRNDA	N	260	ND
CAP-9	137-146	TQDATYLWWV	N	204	ND
CAP-10	102-110	GQFRVYPEL	N	201	ND
NCA-1	571-579	YRPGENLNL	N	252	225
Positive control	—	ALAAAAAAV	P	632	—
No peptide	—	—	—	280	300

*Predicted binding on the basis of published motifs (*31*); P = positive; N = negative.

†Reactivity of T2 cells with anti–HLA-A2 MAb after the cells were incubated with CEA peptide. Peptides were used in a concentration of 50 μg/mL/10^6 cells. The results are expressed in relative fluorescence values (350 was arbitrarily chosen as a cutoff value for positive). Positive control 9-mer is an A2-binding motif.

10^5 pfu (patient Vac7), 2×10^6 pfu (patient Vac6), and 10^7 pfu (patients Vac24, Vac23, and Vac32) per injection. PBMCs from patients were separated using lymphocyte separation medium gradient (Organon Teknika, Durham, NC) as previously described (*29*). Washed PBMCs were resuspended in complete medium: RPMI-1640 (Life Technologies, Inc.) supplemented with 10% pooled human AB serum (Pel Freeze Clinical System), 2 m*M* glutamine, 100 U/mL penicillin, and 100 μg/mL of streptomycin (Life Technologies, Inc.). Cells (2×10^5) in complete medium in a volume of 100 μL were added into each well of a 96-well flat-bottom assay plate (Corning Costar Corp., Cambridge, MA). Peptides were added to cultures at a final concentration of 50 μg/mL. Cultures were incubated for 5 days at 37 °C in a humidified atmosphere containing 5% CO_2. After removal of the peptide-containing medium, the cultures were then provided with human IL-2 (provided by the NCI Surgery Branch) (10 U/mL) for 11 days, with IL-2–containing medium being replenished every 3 days. The incubation time of 5 days with peptide plus 11 days with IL-2 constitutes one cycle. Primary cultures were restimulated with the same peptide (50 μg/mL) on day 16 to begin the next cycle. Irradiated (4000 rad) autologous peripheral blood mononuclear cells (5×10^5) were added in a volume of 50 μL in complete medium as APCs. T-cell lines derived from patients Vac24, Vac6, Vac7, etc. were given the designations V24T, V6T, V7T, etc., respectively.

Cytotoxicity Assays

Various target cells were labeled with 50 μCi of ^{111}In-oxyquinoline (Medi-Physics Inc., Arlington, IL) for 15 minutes at room temperature. Target cells (0.5×10^4) in 100 μL of complete medium (see below) were added to each of 96 wells in U-bottom assay plates (Corning Costar Corp.). The labeled targets were incubated with peptides at a final concentration of 50 μg/mL for 60 minutes at 37 °C in CO_2 before adding effector cells. Effector cells were suspended in 100 μL of complete medium supplemented with 10% pooled human AB serum and added to target cells; the plates were then incubated at 37 °C in 5% CO_2 for 12 or 18 hours. Supernatant was harvested for gamma counting with the use of harvester frames (Skatron, Inc., Sterling, VA). Determinations were carried out in triplicate and standard deviations were calculated. All experiments were carried out three times, except for data in Table 7, which was done twice. Specific lysis was calculated with the use of the following formula:

$$\% \text{ lysis} = \frac{\text{observed release (cpm)} - \text{spontaneous release (cpm)}}{\text{total release (cpm)} - \text{spontaneous release (cpm)}} \times 100.$$

Spontaneous release was determined from wells to which 100 μL of complete medium was added. Total releasable radioactivity was obtained after treatment of targets with 2.5% Triton X-100.

Detection of Tumor Necrosis Factor-α

Supernatants of T cells exposed for 3 days to peptides and APC in IL-2–free medium, at a responder to stimulator ratio of 4:1 (4×10^6:1×10^6 cells/mL), were screened for the secretion of tumor necrosis factor-α (TNF-α), using an enzyme-linked immunosorbent assay kit (Genzyme Corp., Cambridge, MA). The results were expressed in pg/mL.

Flow Cytometry

The procedure for single-color flow cytometric analysis has been previously described (*30*). Briefly, 1×10^6 cells were washed three times with cold Ca^{2+}- and Mg^{2+}-free Dulbecco's phosphate-buffered saline (DPBS) and then stained for 1 hour with 1 μg of monoclonal antibody (MAb) against CD3 (Becton Dickinson, San Jose, CA), CD4 (Becton Dickinson), CD8 (Becton Dickinson), HLA class I (W6/32) (Sera-Lab, Sussex, England), HLA class II (HLA-DR) (Becton Dickinson), and MOPC-21 (Cappel/Organon Teknika Corp., West Chester, PA) in a volume of 100 μL of PBS containing 1% bovine serum albumin. Anti-CEA MAb COL-1 was used as 100 μL of culture supernatant. The cells were then washed three times with cold DPBS and incubated for an additional hour in the presence of 1:100 dilution (volume of 100 μL PBS containing 1% bovine serum albumin) of fluorescein-conjugated goat anti-mouse immunoglobulin (Ig) (Kirkegaard & Perry Labs., Gaithersburg, MD). The cells were again washed three times with DPBS and resuspended in DPBS at a concentration of 1×10^6 cells/mL. The cells were immediately analyzed using a Becton Dickinson FACScan equipped with a blue laser with an excitation of 15 nW at 488 nm. Data were gathered from 10 000 live cells, stored and used to generate results.

The procedure for dual-color flow cytometric analysis was similar to that for single-color analysis, with the following exceptions. The antibodies used were anti-CD4 fluorescein conjugate, anti-CD8 phycoerythrin conjugate, anti-IgG_1 fluorescein conjugate, and anti-IgG_{2a} phycoerythrin conjugate (Becton Dickinson). Staining was done simultaneously for 1 hour after which cells were washed three times, resuspended as above, and immediately analyzed using a Becton Dickinson FACSort equipped with a blue laser with an excitation of 15 nW at 488 nm equipped with the Lysis II program.

Binding of CEA peptides to the HLA-A2 molecule was analyzed by the up-regulation of HLA-A2 expression on T2 cells (*25*) as demonstrated by flow cytometry. The T2 cell peptide binding assay has been reported previously (*31*). Briefly, aliquots of $0.5\text{-}1 \times 10^6$ T2 cells in serum-free IMDM were incubated with peptides at a concentration of 50 μg/mL in 24-well culture plates at 37 °C in 5% CO_2 overnight. Flow cytometry for peptide binding was carried out using T2 cells and single-color analysis. After cells were washed three times in DPBS as above, they were incubated for 1 hour with HLA-A2 specific antibody A2,28 (#189HA-1; One Lambda, Inc., Canoga Park, CA), using 10 μL of a 1× working dilution/10^6 cells. MOPC-104E (Cappel/Organon Teknika Corp., West Chester, PA) was used as isotype control. The cells were then washed three times and incubated with a 1:100 dilution of phycoprobe PE (phycoerythrin) labeled anti-mouse IgM (Biomeda Corp., Foster City, CA). Analysis was carried out using the FACScan as described above. Cells were maintained on ice during all cell preparation and staining unless otherwise stated above.

HLA Typing

The HLA phenotyping of patients was performed by the Tissue Typing QC Laboratory, Naval Medical Research Institute, Bethesda, MD, using a standard antibody-dependent microcytotoxicity assay and a defined panel of anti-HLA antisera. The HLA phenotypes were as follows: patient Vac24 (HLA-A2, 24; B 44, 51; DR 4, 11; DQ 3, 7; DR w52, 53); patient Vac01 (HLA-A28, 31; B14, 35; DR1, 4; DQ1, 3; DRw53); patient Vac23 (HLA-A1, 26; B8, 60; CW3, 7; DR0103, 15; DQ5, 6); patient Vac32 (HLA-A3, 68; B7, 51; CW7; DR4, 15; DQ1, 8; DRw53); patient Vac6 (HLA-A2, 24; B13, 51; CW6; DR7, 8; DQ4; DR53); and patient Vac7 (HLA-A2; B7; CW7; DR15, 17; DQ1, 2; DR52). From patient Vac01, only B cells were used.

Vaccinia Virus Infection of Colorectal Carcinoma Cells

cDNA for the HLA-A2.1 gene in the vaccinia virus vector was provided by the Surgery Branch, NCI, National Institutes of Health. These genes were inserted into the TK gene in plasmid pSCII, allowing homologous recombination to occur with the viral TK gene (*32*). Target cells at a concentration of 1×10^7/mL in complete RPMI-1640 medium supplemented with 0.1% bovine serum albumin were incubated with an equal volume of vaccinia virus (10^9 pfu/mL) in the same medium at 37 °C for 1.5 hours. The cells were then adjusted to a concentration of 5×10^5/mL in complete medium and incubated for 3 hours at 37 °C.

Statistical Analysis

Statistical analysis of differences between means was done by a two-tailed paired *t* test.

Results

Identification of Potential CEA-Specific T-Cell Epitopes

Since the entire amino acid sequence of human CEA is known and human HLA class I-A2 consensus motifs have been described (*33*,*34*), studies were undertaken to identify a series of peptides that would potentially bind class I-A2 molecules. A2 was chosen, since it is the most common HLA class I molecule, being represented in approximately 50% of North American caucasians and 34% of African-Americans (*35*). The peptide sequence of CEA was thus examined for matches to the consensus motifs for HLA-A2 binding peptides. Peptides were only selected, moreover, if their sequence diverged sufficiently from the CEA-related NCA and BGP sequences. The amino acid sequence of human CEA (GeneBank Accession #M17303) was scanned using a predictive algorithm (*36*) that combines a search for anchor residues with numerical assignments to all residues at all positions. Ten peptides were synthesized using this algorithm, ranging in length from 9 to 11 amino acids. Six of these peptides also contained the HLA-A2 binding motif of leucine or isoleucine at position 2 and valine or leucine at the C terminal. Another peptide (CAP-7) also possessed the motif for binding to HLA-A3 (*37*). All peptides were selected to have minimal homology to the parallel regions of NCA and BGP after optimal alignment of the latter sequences with CEA. The 9-mer, 10-mer, or 11-mer peptides that met these criteria were selected for synthesis and purification; they were designated CAP (carcinoembryonic antigen peptide)-1 through 10. Their amino acid sequence and position in the CEA molecule are given in Table 1. The positive (P) or negative (N) designation (Table 1) relates to the predicted binding to HLA-A2.

The T2 cell-binding assay has been used to predict human HLA-A2 consensus motifs (*31*). In this assay, the binding of an appropriate peptide results in the up-regulation of surface HLA-A2 on the T2 cells, which can be quantified via FACScan using an anti–HLA-A2 antibody. As seen in Table 1, six of the CEA peptides (CAP-1 through CAP-6) scored positive for T2 binding (the peptides were designated CAP-1 through CAP-10 retrospectively on the basis of their quantitative binding to T2). In general, the peptides with the HLA-A2 consensus motif were better binders than those lacking the motif. The order of T2 cell-peptide binding did not always correspond to the predictive algorithm (*36*). Since peptide 571-579 (designated CAP-1) demonstrated the highest level of T2 binding, the peptide reflecting the NCA analog (the corresponding NCA peptide obtained after optimal alignment of NCA and CEA) was also synthesized and tested; this peptide, designated NCA-1, showed background binding to T2 cells (Table 1). This low level of binding was consistent with the fact that an amino acid substitution in NCA had abolished one of the A2 anchor residues (Arg for Leu at position 2).

Establishment of T-Cell Lines to CEA Peptides

In an attempt to establish T-cell lines from patients who had received the rV-CEA construct, PBLs were obtained from three patients (designated Vac6, Vac7, and Vac24) with the HLA-A2 allele and were alternately pulsed with 50 μg/mL peptide CAP-1 and IL-2 (10 U/mL) as described in the "Materials and Methods" section. In all three cases, T-cell lines could be established that were cytotoxic for T2 cells when pulsed with the CAP-1 peptide. Fig. 1 shows the results of these assays using T-cell lines from patients Vac24 and Vac6. The T-cell line from patient Vac24 was chosen for further study.

PBLs from patient Vac24 (before and after vaccination with three doses of 10^7 pfu rV-CEA at monthly intervals) were placed in 96-well plates and pulsed with the CAP-1 peptide and then IL-2, as described in the "Materials and Methods" section. Each exposure to peptide and IL-2 was considered one cycle of stimulation. As seen in Table 2, one, two, or three cycles of CAP-1 peptide and IL-2 did not result in growth of cells in any of the 96 wells using the preimmunization PBLs. By contrast, after one cycle of stimulation of post-vaccination PBLs from the same patient, 66 of the 96 wells (69%) demonstrated growth of cells, which were maintained through four cycles of stimulation. It is of interest that, after four cycles of stimulation of preimmunization PBLs, two (2%) of 96 wells exhibited cell growth. Thus, one could hypothesize that a minor population of T cells exists in this patient that is capable of recognizing a specific CEA epitope (peptide 571-579) and that these cells were clonally expanded as a result of the rV-CEA administration.

Sufficient PBLs, before and after vaccination with rV-CEA at the 10^7 pfu dose, were also available from two non–HLA-A2 patients: Vac32 (HLA A1,26) and Vac23 (HLA A3,68). Since we had little or no basis for predicting which peptides might bind to these haplotypes, nine of the CEA peptides were used in an attempt to establish T-cell lines. Using peptide CAP-1 with IL-2 as described above, no T-cell lines could be established from preimmunization PBLs from either patient Vac32 or Vac23 (Table 2). However, employing post–rV-CEA immunization PBLs, T-cell lines were established after three cycles of stimulation in 25 (52%) of 48 wells for patient Vac32 and in 21 (44%) of 48 wells for patient Vac23 (Table 2).

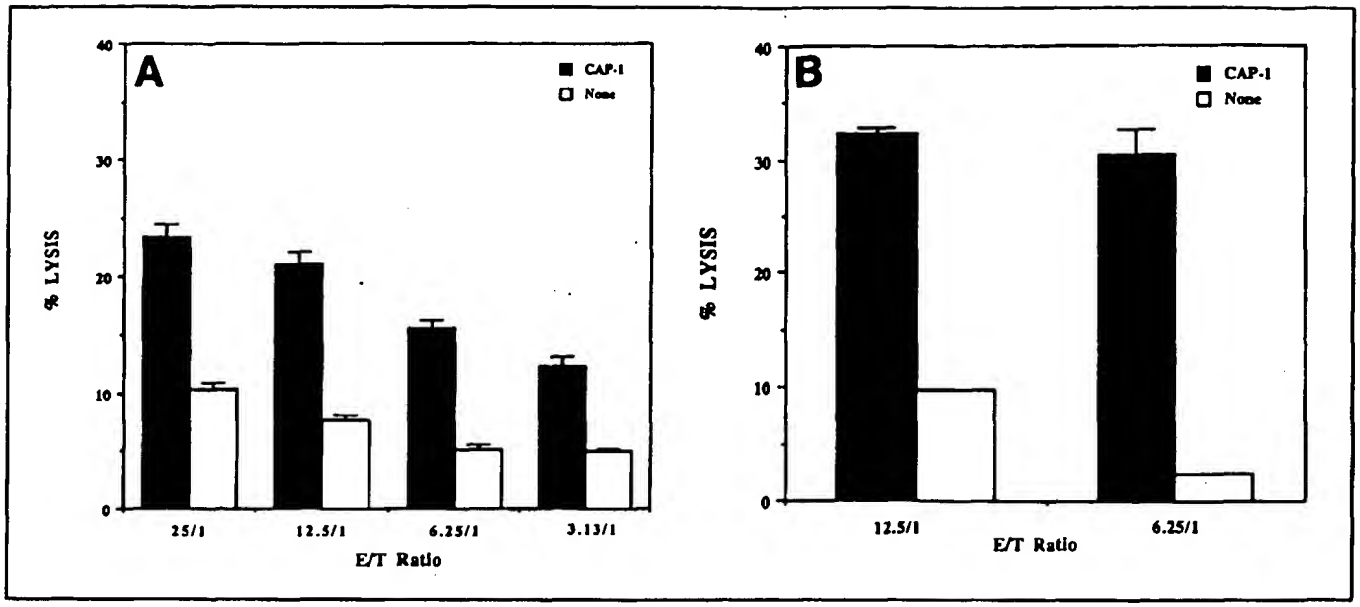


Fig. 1. Cytotoxicity of T-cell lines (designated V24T [Panel A] and V6T [Panel B]) derived from patients immunized with rV-CEA and induced by CEA CAP-1 peptide. CTL activity was determined in an 18-hour ^{111}In release assay using T2 cells as a target incubated with CAP-1 peptide (50 μg/mL). E/T ratio = effector-to-target ratio.

A similar contrast in pre-vaccination versus post-vaccination PBLs from patients Vac32 and Vac23 was seen with a mixture of CEA peptides CAP-4, CAP-6, and CAP-7 (Table 2). Combinations of peptides were used to conserve PBLs. It is of interest that PBL from patient Vac32 (HLA-A3 positive) showed evidence of cell growth in the presence of CAP-7, the peptide that bears the HLA-A3–binding motif. It should be noted that these results suggest that a peptide shown to bind to HLA-A2 can also stimulate T-cell lines after binding to some non-A2 antigens; the possible reasons for this will be discussed in detail below. However, it was decided to first characterize the T-cell responses in patient Vac24 because of the implied relevance of MHC binding and T-cell activation. Nonetheless, it is encouraging that PBLs from five of five patients showed signs of T-cell response to peptide CAP-1 after immunization with rV-CEA.

Flow Cytometry Analysis

Flow cytometric studies were conducted to phenotype the V24T cell line and V6T and V7T cells obtained by pooling respective cells growing in 96-well plates. The results are shown in Table 3. Cells stained double positive for both CD8 and CD4 in V24T and V6T cell lines, while the V7T cell line was CD8 positive.

Cytotoxicity Assays

To determine if the T-cell line from patient Vac24 (designated V24T) could lyse autologous B cells presenting the CAP-1 peptide, B cells from this patient were first transformed with EBV and then pulsed (i.e., incubated) with the CAP-1 peptide. As seen in Table 4, the V24T cells were capable of lysing the

Table 2. Frequency of T-cell growth following an in vitro stimulation with CEA peptides: before and after immunization with rV-CEA

		Frequency							
		Before immunization*				After immunization*			
Patient	Peptide	1	2	3	4	1	2	3	4
Vac24†	CAP-1	0	0	0	2	66	66	66	66
Vac32‡	CAP-4, CAP-6, CAP-7	0	0	4	ND	0	0	30	ND
	CAP-3, CAP-10, CAP-9	0	0	0	ND	0	0	0	ND
	CAP-8, CAP-2	0	0	0	ND	0	0	0	ND
	CAP-1	0	0	0	ND	0	0	25	ND
Vac23‡	CAP-4, CAP-6, CAP-7	0	0	3	ND	0	0	31	ND
	CAP-3, CAP-10, CAP-9	0	0	0	ND	0	0	0	ND
	CAP-8, CAP-2	0	0	0	ND	0	0	0	ND
	CAP-1	0	0	0	ND	0	0	21	ND

*Cycle of stimulation with peptide and IL-2 (*see* "Materials and Methods" section); ND = not done.

†Results are expressed as the number of positive wells per 96 wells. Peripheral blood mononuclear cells were seeded at a concentration of 1×10^5 cells per 200 μL/well.

‡Results are expressed as the number of positive wells per 48 wells. Peripheral blood mononuclear cells were seeded at a concentration of 1×10^6 cells per well in 1 mL.

Table 3. Flow cytometric analysis of surface markers on T-cell lines*

Surface antigen†	Cell line V24T	V6T	V7T
$CD8^+/CD4^+$	62.4	30.8	Negative
$CD8^+/CD4^-$	35.7	56.0	99.6
$CD4^+/CD8^-$	Negative	12.6	Negative
$CD3^+$	98.5	95.4	99.4

*Negative = less than 5% positive. Results are expressed in percentage of each T-cell line reactive with the MAbs. Routinely, 2%–4% of the cells were stained when treated either with no primary MAb or an isotype-matched control MAb.
†+ = positive; – = negative.

Table 4. Ability of V24T cells (T cells derived from PBLs from rV-CEA–vaccinated patient Vac24) to lyse autologous and allogeneic B cells pulsed with a CEA-specific peptide*

B cells	HLA-A2†	Pulsing peptide	% lysis (± SD)
B-Vac24	Positive	CAP-1	43 (0.29)†
		NCA-1	8 (1.92)
		None	2 (0.37)
B-Vac01	Negative	CAP-1	10 (0.53)
		NCA-1	11 (0.62)†
		None	8 (0.68)

*An 18-hour ^{111}In release assay was performed. Peptides were used in a concentration of 50 μg/mL. Results are expressed as percentage specific lysis at effector-to-target ratios of 25:1.
†Statistically significant ($P<.01$, paired t test). Similar statistically significant lysis was observed at effector-to-target ratios of 12.5:1.

autologous B cells when pulsed with CAP-1, but when an allogeneic (non–HLA-A2) EBV-transformed B cell was pulsed with the same peptide, no lysis was observed. Lysis was observed at effector-to-target cell ratios of 25:1 and 12.5:1. When the NCA-1 peptide, reflecting the analogous region on the NCA molecule was used to pulse B cells of patient Vac24, no lysis was observed with the V24T cells. As shown in Table 1, this lack of lysis was not unexpected, since three of the nine amino acids of NCA-1 differ from those of CAP-1, including an anchor residue.

Studies were then undertaken to determine if the CAP-1 peptide could induce the secretion of TNF-α from the cytolytic V24T cells. Incubation of V24T cells with autologous B cells pulsed with CAP-1 peptide resulted in the production of more than 300 pg/mL of TNF-α, while incubation with control or non-T2 cell-binding peptides CAP-9 and CAP-10 or no peptide showed levels of production below 75 pg/mL.

Cytotoxicity of V24T Cells Against Tumor Cells

While the above studies indicate that autologous B cells can present the CAP-1 peptide to the V24T cells, resulting in lysis of the B cells, they do not indicate that human APCs can endogenously process the entire CEA molecule in a manner so as to bind HLA-A2 molecules for presentation at the cell surface. To help clarify this issue, EBV-transformed B cells of patient Vac24 were transduced with the entire human CEA gene, using a retroviral vector (*see* "Materials and Methods" section). As seen in Table 5, the CEA-transduced cells now express CEA, and the transduction process had no effect on the expression of HLA class I and class II molecules.

Table 5. Flow cytometric analysis of surface antigens of EBV-transformed B cells derived from patient Vac24 before and after transfection with CEA

Antigen	MAb	% positive* B-Vac24†	B-Vac24 (CEA)‡
CEA	COL-1	4.4 (24.7)	42.9 (100.9)
HLA class I	W6/32	100.0 (831.9)	100.0 (519.1)
HLA class II	anti–HLA-DR	99.8 (313.3)	99.3 (221.8)
Control	MOPC-21	2.2 (30.5)	2.0 (24.1)

*Values represent the percentage of each cell type reactive with MAbs listed as analyzed by flow cytometry. Numbers in parentheses are the mean channel fluorescence intensity as determined in relative log units. Routinely, 2%–4% of the cells were stained when treated either with no primary MAb or an isotype-matched control MAb. LS-174T, a colorectal carcinoma cell line, was used as a positive control for CEA expression. The percent positive value for CEA in CEA-expressing LS-174T human colon carcinoma cell line was 59.8 (144.4).
†B-Vac24 are EBV-transformed B cells derived from PBLs of patient Vac24 prior to immunization with rV-CEA.
‡B-Vac24(CEA) are the same as B-Vac24 except they have been transduced with the entire human CEA gene using a retroviral vector as described in the "Materials and Methods" section.

As shown in Table 6, the autologous B cells transduced with the CEA gene can now serve as targets for the V24 CTLs. These results thus demonstrate that a CEA gene product can be endogenously processed by autologous B cells and presented at the cell surface in the context with class I MHC to induce T-cell lysis. The question now remained as to whether human carcinoma cells can act in the same manner as APCs and, thus, serve as potential targets for V24T cells. As seen in Table 6, non-A2 allogeneic carcinoma cells SW1417 and HT-29, which do express substantial CEA, cannot serve as targets, while the allogeneic A2-positive SW403 carcinoma cells expressing CEA are lysed at effector-to-target ratios of 50:1 and 25:1.

Cytotoxicity of V24T Cells Against Vaccinia-CEA–Infected Tumor Cells

To further demonstrate the HLA-A2 restricted nature of the V24T cells in the lysis of human carcinoma cells, the CEA-positive, non–HLA-A2 SW837 human carcinoma cell line was employed. These cells were either uninfected, infected with wild-type vaccinia virus, or infected with a recombinant vaccinia virus containing the HLA-A2 gene. Twelve-hour lysis ex-

Table 6. Cytotoxicity of V24T cell line (derived from patient Vac24 immunized with rV-CEA) on target cells with endogenous CEA expression*

Target	HLA-A2	CEA	% lysis (± SD)
B-Vac24†	Positive	Negative	8.2 (2.1)
B-Vac24 (CEA)†	Positive	Positive	46.1 (11.6)§
SW403‡	Positive	Positive	45.2 (1.5)§
SW1417‡	Negative	Positive	5.2 (0.5)
HT-29‡	Negative	Positive	4.1 (0.6)

*HLA-A2 and CEA expression were tested by flow cytometry using MAbs anti-A2 and COL-1, respectively. An 18-hour ^{111}In release assay was performed. Results are expressed in percent specific lysis at effector-to-target ratio of 50:1 compared with lysis obtained with B-Vac24 cells. Similar statistically significant lysis was seen at effector-to-target ratio of 25:1.
†As described in the legend to Table 4.
‡Human colon carcinoma cell lines expressing CEA.
§Statistically significant lysis ($P<.01$, paired t test).

periments were carried out to avoid spontaneous lysis due to vaccinia virus. As seen in Table 7, only the carcinoma cells infected with the rV-A2.1 recombinant expressing HLA-A2 were susceptible to lysis with V24T cells. These studies further demonstrate the HLA-A2 restricted nature of the CEA-specific lysis of the V24T cells.

Discussion

These studies demonstrate that one can evoke a cytotoxic T-cell response to a specific epitope of the human CEA molecule by vaccination with rV-CEA. This response appears to be mediated via a class I MHC restricted mechanism. Moreover, the lytic T cells generated against the defined CEA peptide were able to lyse tumor cells endogenously synthesizing the entire CEA antigen. The ability to raise a CTL response versus CEA in humans was clearly not a foregone conclusion prior to these studies; since CEA is expressed in fetal tissue and some normal colonic mucosa, "tolerance" to this molecule was one possible outcome. Moreover, previous studies dealing with the presence of antibodies to CEA were inconclusive and contradictory (*38-41*), and no studies have reported either the presence or the absence of cytolytic T-cell responses to CEA in humans. It was for this reason that the CEA gene was placed into vaccinia virus. A previous study (*13*) in a mouse model demonstrated enhanced CTL responses when employing rV-CEA as an immunogen as opposed to native CEA.

Other data are currently being accumulated on the potential immunogenicity of CEA. While the patients in the study reported here had advanced metastatic disease, a trial using the same rV-CEA construct described here, at the same dose schedule and route of administration, is currently under way in gastrointestinal cancer patients with minimal disease. In those studies, primary lymphoproliferative T-cell responses to CEA have been observed (*42*). In another study (*43*), an anti-idiotype MAb to an anti-CEA MAb has been administered to gastrointestinal cancer patients; in that study, antibodies to CEA and lymphoproliferative responses to CEA were reported. It should also be noted that a pilot phase I study (*44*) was previously carried out in which soluble CEA was administered to patients with advanced gastrointestinal carcinomas. In that study, as well as the other studies cited, no toxicity was observed. The findings of these phase I trials thus indicate that CEA deserves further evaluation as a potential target for immunotherapeutic applications.

Table 7. Demonstration of HLA-A2 involvement in ability of V24T cells to lyse human colon carcinoma cells*

Infection	Expression of MHC class I molecules†		% lysis (± SD)‡
	W6/32 antibody	A2,28 antibody	
None	99.1 (228.23)	4.4 (47.41)	8.2 (1.4)
Vaccinia A2.1	99.3 (261.74)	90.4 (66.70)	31.4 (2.5)§
Vaccinia vector control	99.4 (217.59)	2.9 (17.54)	7.2 (0.6)

*SW837, an HLA-A2–negative and CEA-positive human colorectal carcinoma cell line, was infected with vaccinia A2.1 vector. The surface expression of HLA-A2 molecule after infection was analyzed by flow cytometry using anti–HLA-A2 MAb (A2.28) and anti-HLA class I antibody (W6/32).

†Values represent the percentage of cells reactive with the antibodies listed. Numbers in parentheses are the mean fluorescence intensity as determined in relative log units.

‡Results are expressed in percent specific lysis in a 12-hour ^{111}In release assay at effector-to-target ratio of 50:1.

§Statistically significant lysis ($P<.01$, paired t test).

The finding reported here that the administration of rV-CEA can elicit a T-cell response to a defined 9-mer epitope on CEA leads to potential second-generation protocols. One of the limitations on the use of a vaccinia recombinant vaccine is that it elicits a strong antibody response to the vaccinia virus (*13*). This in turn, prohibits numerous boosts with the recombinant vaccinia, due to the anti-vaccinia antibody inhibition of local virus spread. Indeed, this boost in anti-vaccinia antibody, seen in the phase I trial with rV-CEA, may be a limitation in the sole use of rV-CEA as immunogen.

Now that a specific epitope on the CEA molecule has been identified, subsequent studies in HLA-A2 patients can be designed in which one or two rV-CEA administrations are followed by several administrations of the CAP-1 peptide (amino acid 571-579) in adjuvant or liposomes as boosts. Moreover, protocols involving the use of rV-CEA and recombinant CEA protein, anti-idiotype MAb (as surrogate antigen for CEA), or other viral vectors may be considered. It should be pointed out that the T cells recognizing the CAP-1 peptide do not recognize the corresponding sequence on NCA. In addition, the specificity of V24T was demonstrated by the induction of TNF-α secretion by incubation of V24T with autologous B cells pulsed with CAP-1 but not with CAP-9 or CAP-10 peptides.

An alternative approach would be to expand in vitro the CTL lines derived from PBLs of rV-CEA–immunized HLA-A2 patients by pulsing with the CAP-1 peptide plus IL-2. These antigen-specific CTLs could then be adoptively transferred to the autologous patient. The adoptive transfer of antigen-specific CTL could conceivably be followed by administration of the CAP-1 peptide in adjuvant or a liposomal formulation to further expand the CEA-specific CTL population in vivo. Perhaps fewer cells of this type need to be adoptively transferred than in tumor-infiltrating lymphocyte protocols (*45*), since these antigen specific T cells are expanded and activated by a known 9-mer epitope.

The broad use of peptide CAP-1 as an immunogen is an intriguing issue. We saw evidence of T-cell growth and the establishment of CTL lines from post-immunization PBLs from three of three HLA-A2 individuals. Unexpectedly, we also saw indications of cell growth in post-immunization PBLs from two individuals who were negative for the A2 allele. In both of these cases, however, T-cell growth was not seen until after three cycles of stimulation with peptide and IL-2. If in fact these are MHC-restricted T-cell responses, one must postulate that the peptide designed to bind to HLA-A2 molecules can also bind to other HLA class I or II molecules. In principle, this phenomenon could arise from the following scenarios: (*a*) alternate class I or II molecules share consensus motifs with HLA-A2 or (*b*) the CAP-1 peptide fortuitously contains sequences capable of binding more than one HLA class I or II antigen, i.e., contains multiple consensus motifs. The current literature on MHC restriction offers examples of both scenarios. For example, HLA-

A3 and HLA-A11 molecules are both known to prefer 9-mer peptides with valine at position 2 and lysine at the C terminal (*46*). An example of the second situation is the specific immunity generated in mice of different genetic backgrounds to a defined viral epitope of lymphocytic choriomeningitis virus (*47*).

Less controversial, but equally intriguing, is the suggestion that T-cell growth was observed in vitro in HLA-A3 PBL (patient Vac32) by stimulation with a peptide that also fits the A3 consensus motif. If this HLA-A3 stimulation turns out to be the case, evidence is provided that rV-CEA can induce MHC-restricted CTL responses in individuals with various haplotypes.

The area of T-cell immunotherapy that is now emerging involves at least two vital parts. The first is the identification of the target antigen/epitope in the context of known MHC molecules in terms of recognition by T cells. The second vital step involves T-cell activation and clonal expansion. Thus, studies are now under way to better activate T cells recognizing specific human epitopes. These studies will involve the use of cytokines and T-cell costimulatory molecules such as B7.1. In a recent study (*48*), we have demonstrated that the murine B7.1 and B7.2 molecules can each be placed into a vaccinia virus vector to enhance the immunogenicity of murine carcinoma cells in an experimental model. Studies are currently under way in an experimental model to coadminister rV-CEA with rV-B7 recombinants to enhance the immunogenicity of CEA; preliminary results appear to be promising.

References

(*1*) van der Bruggen P, Traversari C, Chomez P, et al: A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science 254:1643-1647, 1991

(*2*) Horn SS, Topalian SL, Simonis T, et al: Common expression of melanoma tumor-associated antigens recognized by human tumor infiltrating lymphocytes: analysis by human lymphocyte antigen restriction. J Immunother 10:153-164, 1991

(*3*) Cox AL, Skipper J, Chen Y, et al: Identification of a peptide recognized by five melanoma-specific human cytotoxic T cell lines. Science 264:716-719, 1994

(*4*) Darrow TL, Slingluff CL Jr, Seigler HF: The role of HLA Class I antigens in recognition of melanoma cells by tumor-specific cytotoxic T lymphocytes. Evidence for shared tumor antigens. J Immunol 142:3329-3335, 1989

(*5*) Oesterling JE: Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. J Urol 145:907-923, 1991

(*6*) Peace DJ, Xue BH, Sosman JA, et al: In vitro immunization of human cytotoxic T lymphocytes specific for peptides derived from prostate specific antigen. Cancer Vaccine Symposium. Cancer Research Institute. Oct. 3-5, 1994 [abstr]

(*7*) Fisk B, Chesak B, Ioannides MG, et al: Sequence motifs of human HER-2 proto-oncogene important for peptide binding to HLA-A2. Int J Oncol 5:51-63, 1994

(*8*) Ioannides CG, Fish B, Jerome KR, et al: Cytotoxic T cells from ovarian malignant tumors can recognize polymorphic epithelial mucin core peptides. J Immunol 151:3693-3703, 1993

(*9*) Tsang KY, Nieroda CA, De Filippi R, et al: Induction of human cytotoxic T-cell lines directed against point-mutated p21 ras-derived synthetic peptides. Vaccine Res 3:183-193, 1994

(*10*) Jung S, Schluesener HJ: Human T lymphocytes recognize a peptide of single point-mutated, oncogenic ras proteins. J Exp Med 173:273-276, 1991

(*11*) Fenton RG, Taub DD, Kwak LW, et al: Cytotoxic T-cell response and in vivo protection against tumor cells harboring activated ras proto-oncogenes [*see* comment citation in Medline]. J Natl Cancer Inst 85:1294-1302, 1993

(*12*) Houbiers JG, Nijman HW, van der Burg SH, et al: In vitro induction of human cytotoxic T lymphocyte responses against peptides of mutant and wild-type p53. Eur J Immunol 23:2072-2077, 1993

(*13*) Kantor J, Irvine K, Abrams S, et al: Antitumor activity and immune responses induced by a recombinant carcinoembryonic antigen-vaccinia virus vaccine [*see* comment citation in Medline]. J Natl Cancer Inst 84:1084-1091, 1992

(*14*) Kantor J, Irvine K, Abrams S, et al: Immunogenicity and safety of a recombinant vaccinia virus vaccine expressing the carcinoembryonic antigen gene in a nonhuman primate. Cancer Res 52:6917-6925, 1992

(*15*) Ras E, VanderBerg SH, Warnaar SO: Identification of potential tumor-specific CTL epitopes of epithelial surface antigen (ESA) and carcinoembryonic antigen (CEA). European Immunology Meeting, Barcelona, June 1994 [abstr]

(*16*) Thompson JA, Grunert F, Zimmermann W: Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. J Clin Lab Anal 5:344-366, 1991

(*17*) Oikawa S, Imajo S, Noguchi T, et al: The carcinoembryonic antigen (CEA) contains multiple immunoglobulin-like domains. Biochem Biophys Res Commun 144:634-642, 1987

(*18*) Staab HJ, Anderer FA, Stumpf E, et al: Are circulating CEA immune complexes a prognostic marker in patients with carcinoma of the gastrointestinal tract? Br J Cancer 42:26-33, 1980

(*19*) Mavligit GM, Stuckey S: Colorectal carcinoma. Evidence for circulating CEA-anti-CEA complexes. Cancer 52:146-149, 1983

(*20*) Collatz E, Von Kleist S, Burtin P: Further investigations of circulating antibodies in colon cancer patients: on the autoantigenicity of the carcinoembryonic antigen. Int J Cancer 8:298-303, 1971

(*21*) Chester KA, Begent RH: Circulating immune complexes (CIC), carcinoembryonic antigen (CEA) and CIC containing CEA as markers for colorectal cancer. Clin Exp Immunol 58:685-693, 1984

(*22*) Ura Y, Ochi Y, Hamazu M, et al: Studies on circulating antibody against carcinoembryonic antigen (CEA) and CEA-like antigen in cancer patients. Cancer Lett 24:283-295, 1985

(*23*) Hamilton JM, Chen AP, Nguyen B, et al: Phase I study of recombinant vacinnia virus (rV) that expresses human carcinoembryonic antigen (CEA) in adult patients with adenocarcinomas. Proc ASCO:abstr 961, 1994

(*24*) Gritz L, Destree A, Cormier N, et al: Generation of hybrid genes and proteins by vaccinia virus-mediated recombination: application to human immunodeficiency virus type I env. J Virol 64:5948-5957, 1990

(*25*) Anderson KS, Alexander J, Wei M, et al: Intracellular transport of class I MHC molecules in antigen processing mutant cell lines. J Immunol 151:3407-3419, 1993

(*26*) Blumberg RS, Panadis T, Byington R, et al: Effects of human immunodeficiency virus on the cellular immune response to Epstein-Barr virus in homosexual men: characterization of the cytotoxic response and lymphokine production. J Infect Dis 155:877-890, 1987

(*27*) Robbins PF, Kantor JA, Salgaller M, et al: Transduction and expression of the human carcinoembryonic antigen gene in a murine colon carcinoma cell line. Cancer Res 51:3657-3662, 1991

(*28*) Tsang KY, Kashmiri SV, De Filippi R, et al: A human T cell line engineered to secrete chimeric monoclonal antibody. J Immunother 13:143-152, 1993

(*29*) Boyum A: A one-stage procedure for isolation of granulocytes and lymphocytes from human blood. General sedimentation properties of white blood cells in a 1g gravity field. Scand J Clin Lab Invest 97(suppl):51-76, 1968

(*30*) Guadagni F, Witt PL, Robbins PF, et al: Regulation of carcinoembryonic antigen expression in different human colorectal tumor cells by interferon-γ. Cancer Res 50:6248-6255, 1990

(*31*) Nijman HW, Houbiers JG, Vierboom MP, et al: Identification of peptide sequences that potentially trigger HLA-A2.1-restricted cytotoxic T lymphocytes. Eur J Immunol 23:1215-1219, 1993

(*32*) O'Neil BH, Kawakami Y, Restifo NP, et al: Detection of shared MHC-restricted human melanoma antigens after vaccinia virus-mediated transduction of genes coding for HLA. J Immunol 151:1410-1418, 1993

(*33*) Falk K, Rotzschke O, Stevanovic S, et al: Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290-296, 1991

(*34*) Hunt DF, Henderson RA, Shabanowitz J, et al: Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 255:1261-1263, 1992

(*35*) Lee J: The HLA System: A New Approach. New York: Springer-Verlag, 1990, p 154

(*36*) Parker KC, Bednarek MA, Coligan JE: Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152:163-175, 1994

(*37*) DiBrino M, Parker KC, Shiloach J, et al: Endogenous peptides bound to HLA-A3 possess a specific combination of anchor residues that permit identification of potential antigenic peptides. Proc Natl Acad Sci U S A 90:1508-1512, 1993

(38) Fuchs C, Krapf F, Kern P, et al: CEA-containing immune complexes in sera of patients with colorectal and breast cancer-analysis of complexed immunoglobulin classes. Cancer Immunol Immunother 26:180-184, 1988
(39) Abeyounis CJ, Wilhelm SA, Rittenhouse KW, et al: Is carcinoembryonic antigen an autoantigen? Transplant Proc 16:489-491, 1984
(40) MacSween JM: The antigenicity of carcinoembryonic antigen in man. Int J Cancer 15:246-252, 1975
(41) Steele G Jr, Lahey S, Rodrick M, et al: Circulating immune complexes in patients with colorectal cancer. Am J Surg 145:549-553, 1983
(42) Conry RM, Saleh MN, Schlom J, et al: Breaking tolerance to carcinoembryonic antigen with a recombinant vaccinia virus vaccine in man. Proc Am Assoc Cancer Res 36:abstr 492, 1995
(43) Foon KA, Chakrabarty M, John W, et al: Active immunity to the carcinoembyonic antigen (CEA) in patients treated with an anti-idiotype monoclonal antibody vaccine. Proceedings of the 9th Annual Meeting of the Society for Biological Therapy, Oct 26-20, 1994
(44) Fishbein GE, McClay E, Berd D, et al: A postsurgical adjuvant trial of CEA immunization in patients with Dukes C and D colorectal cancer: a pilot study. Vaccine Res 1:123-128, 1992
(45) Rosenberg SA, Yannelli JR, Yang JC, et al: Treatment of patients with metastatic melanoma with autologous tumor-infiltrating lymphocytes and interleukin 2. J Natl Cancer Inst 86:1159-1166, 1994
(46) Gavioli R, Kurilla MG, DeCampos-Lima PO, et al: Multiple HLA-A11-restricted cytotoxic T-lymphocyte epitopes of different imunogenicities in the Epstein-Barr virus-encoded nuclear antigen 4. J Virol 67:1572-1578, 1993
(47) Oldstone MB, Tishon A, Geckeler R, et al: A common antiviral cytotoxic T-lymphocyte epitope for diverse major histocompatibility complex haplotypes: implications for vaccination. Proc Natl Acad Sci U S A 89:2752-2755, 1992
(48) Hodge JW, Abrams S, Schlom J, et al: Induction of antitumor immunity by recombinant vaccinia viruses expressing B7-1 or B7-2 costimulatory molecules. Cancer Res 54:5552-5555, 1994

Notes

We thank Mr. Marion Taylor, Dr. Jennifer Ng, and Dr. Yoomie Chung for cell cultures and HLA typing, and Dr. Kenneth Parker for assistance in the determination of HLA motifs. We also thank Therion Biologics Corporation for supplying the rV-CEA vaccine, and Ms. Kathleen Siler for her assistance in the preparation of this manuscript.

Manuscript received December 20, 1994; revised March 14, 1995; accepted April 12, 1995.

Int. J. Cancer: 73, 125–130 (1997)

Publication of the International Union Against Cancer
Publication de l'Union Internationale Contre le Cancer

IDENTIFICATION OF HLA-A*0201-RESTRICTED CTL EPITOPES ENCODED BY THE TUMOR-SPECIFIC *MAGE*-2 GENE PRODUCT

Marjan J.W. VISSEREN[1], Sjoerd H. VAN DER BURG[1], Ellen I.H. VAN DER VOORT[1], Remco M.P. BRANDT[1], Peter I. SCHRIER[2], Pierre VAN DER BRUGGEN[3], Thierry BOON[3], Cornelis J.M. MELIEF[1]* and W. Martin KAST[4]

[1]*Department of Immuno-hematology and Blood Bank, Leiden University Hospital, the Netherlands*
[2]*Department of Clinical Oncology, Leiden University Hospital, the Netherlands*
[3]*Ludwig Institute for Cancer Research, Brussels, Belgium*
[4]*Cardinal Bernardin Cancer Center, Loyola University of Chicago, Maywood, IL, USA*

MAGE-2 is expressed in many tumors, including melanoma, laryngeal tumors, lung tumors and sarcomas, but not in healthy tissue, with the exception of testis. Thus, _MAGE_-2-derived peptides that bind to HLA class I molecules and elicit cytotoxic T lymphocyte (CTL) responses could be of significant therapeutic importance. In this study, we show that several _MAGE_-2-derived peptides bind with high affinity to HLA-A*0201. Three of them form complexes with HLA-A*0201 that are stable at 37°C and are immunogenic in HLA-A*0201K^b transgenic mice. Moreover, CTLs against 2 of them (M2 112-120, and M2 157-166) specifically recognize cells that express both the _MAGE_-2 protein and HLA-A*0201K^b. These 2 peptides are processed and presented in the context of HLA-A*0201. Therefore, these peptides are candidate components in peptide-based vaccines for the treatment and prevention of several types of _MAGE_-2-expressing cancers. *Int. J. Cancer* 73:125–130, 1997.
© *1997 Wiley-Liss, Inc.*

Cytotoxic T lymphocyte (CTL) responses are important in the defense against viruses and tumors (Melief and Kast, 1992). CTLs recognize small peptides that are derived from cellular proteins and are presented by MHC class I molecules at the cell surface (Townsend *et al.*, 1986). Tumor cells do express tumor-specific proteins that can potentially be recognized by CTLs capable of eliminating the tumor cells, but such a CTL response is not always elicited for a number of reasons. Tumor cells can use several mechanisms to evade a CTL response, such as down-modulation of MHC class I molecules (Ruiter *et al.*, 1984), secretion of cytokines that frustrate CTL activation (Torre-Amione *et al.*, 1990) and down-modulation of the CTL ζ-chain (Mizoguchi *et al.*, 1992). Suboptimal expression of tumor-specific peptides also precludes induction of a CTL response. This can be circumvented by forced induction of a CTL-mediated anti-tumor response, for example by vaccination with killed tumor cells or tumor specific peptides, or by other modes of specific vaccination, either with or without added cytokines. Once insufficient immunogenicity of tumor cells has been overcome by vaccination, effective CTL-mediated immunity can be demonstrated (Visseren *et al.*, 1994).

Peptide-based vaccines preferably contain peptides derived from tumor-specific proteins that are processed and presented by MHC class I molecules on the cell surface of the tumor cells. One example of a possibly useful tumor-specific protein is the *MAGE*-2 gene product. It is expressed in 70% of metastatic melanomas but also in other types of tumors such as 46% of laryngeal tumors and 35% of lung tumors and sarcomas (De Smet *et al.*, 1994). Moreover, normal fetal and adult tissues do not express it, with the exception of testis tissue (De Plaen *et al.*, 1994). Furthermore, other members of the *MAGE* gene family, most of which are expressed less frequently, have been shown to give rise to peptides that are processed and presented in MHC class I (Traversari *et al.*, 1992). It is therefore conceivable that *MAGE*-2 derived peptides are also presented in HLA class I molecules. Since HLA-A*0201 is a very common HLA type (Imanishi *et al.*, 1992), a vaccine based on *MAGE*-2-derived HLA-A*0201 binding peptides could be of therapeutic importance.

Immunogenicity of HLA-A*0201 binding peptides, and their natural presence in cells expressing both antigen and appropriate HLA, can be studied in HLA-A*0201K^b transgenic mice (Vitiello *et al.*, 1991). These mice express the chimeric HLA-A*0201K^b molecule that consists of the human α1 and α2 domains of HLA-A*0201, forming the peptide-binding groove, and the α3 domain of H2-K^b, allowing binding of murine CD8. This chimeric molecule enables murine CD8$^+$ CTL to interact in a physiological way with HLA-A*0201-peptide complexes. Murine peptide-specific CTL can subsequently be tested on cells expressing both HLA-A*0201K^b and antigen, to study natural processing of the peptides investigated.

In this study, we have identified 7 *MAGE*-2-derived peptides that bind with sufficient affinity to HLA-A*0201. Three of them are able to form stable complexes with this MHC class I molecule at physiological temperature, which is an important feature *in vivo*. These peptides were capable of eliciting a CTL response in HLA-A*0201K^b transgenic mice. Furthermore, we observed that at least 2 of these peptides are processed and presented by HLA-A*0201. Therefore, these are candidate peptides in vaccines for treatment and prevention of several types of *MAGE*-2-expressing cancers.

MATERIAL AND METHODS

Cell culture

The EBV-transformed B cell line JY, homozygous for HLA-A*0201, was cultured in RPMI 1640 (Life Technologies, Paisley, UK) supplemented with penicillin (100 IU/ml, Brocades Pharma, Leiderdorp, the Netherlands), L-glutamine (2 mM, ICN Biochemicals, Costa Mesa, CA), and 8% heat-inactivated FCS (Hyclone, Logan, UT). Jurkat A*0201K^b cells are derived from the human T cell line Jurkat upon transfection with the chimeric *HLA-A*0201K^b* gene (Vitiello *et al.*, 1991). Jurkat A*0201K^b and COS-7 cells were cultured in Iscove's modified Dulbecco's medium with Glutamax I (IMDM; Life Technologies) supplemented with penicillin, 8% heat-inactivated FCS and 20 μM 2-ME (Merck, Darmstadt, Germany). WEHI 164 clone 13 cells were cultured in RPMI 1640 supplemented with 8% heat-inactivated FCS, penicillin, L-glutamine (216 mg/ml), L-asparagine (36 mg/ml) and L-arginine-HCl (116 mg/ml). The *MAGE*-3-specific CTL clone LB705-CTL 297/22 (van der Bruggen *et al.*, 1994) was cultured using standard CTL culture procedures.

Synthetic peptides

Peptides were made by Fmoc chemistry with the multiple peptide synthesizer AMS 422 (Abimed, Langenfeld, Germany). Upon synthesis peptides were analyzed by reversed phase HPLC,

Contract grant sponsor: Macropa Stichting; Contract grant sponsor: Dutch Cancer Society; Contract grant number: NKB 91-15.

*Correspondence to: Department of Immunohematology and Blood Bank, Building 1, E3-Q, Leiden University Hospital, PO Box 9600, 2300 RC Leiden, the Netherlands. Fax: +31 5216751. E-mail: ihbsecr@euronet.nl

Received 6 March 1997; Revised 24 May 1997

which showed over 90% purity. Peptides were lyophilized, weighed, dissolved in DMSO (Merck) at 100 mg/ml and stored at −70°C. From this stock solution peptide was dissolved in PBS at 1 mg/ml and adjusted to pH 7 prior to use in peptide binding, stability, immunogenicity or cytotoxicity experiments. Fluorescein-labeled peptide was synthesized as described previously (van der Burg *et al.*, 1996). Briefly, labeling of a cysteine was performed with 4-(iodoacetamido)-fluorescein (Fluka, Buchs, Switzerland). Fluorescein-labeled peptide was purified by reversed phase HPLC.

Peptide binding assay

Affinity of peptides for HLA-A*0201 was measured as described by van der Burg *et al.* (1996). Briefly, peptides were stripped off the HLA-A*0201-peptide complexes on JY cells, by exposing them for 90 sec to ice-cold citric acid buffer with a final pH of 3.2 (mixture of an equal volume of 0.263 M citric acid and 0.123 M Na_2HPO_4). Subsequently, cells were washed twice with IMDM and resuspended in IMDM supplemented with 1.5 μg/ml $\beta_2$microglobulin (Sigma, St. Louis, MO). In a 96-well U-bottomed plate 150 nM of the fluorescein-labeled reference peptide FLPSD-C(fl)FPSV and titrated amounts of competitor peptide were incubated with 7×10^5 acid-stripped JY cells for 24 hr at 4°C. Cells were washed with PBS containing 1% BSA, fixed with 0.5% paraformaldehyde and analyzed on FACScan (Becton Dickinson, Mountain View, CA). The percentage inhibition was calculated using the following formula:

$$\left(1 - \frac{\text{MF experimental well} - \text{MF background}}{\text{MF reference peptide only} - \text{MF background}}\right) \times 100\%$$

where mean fluorescence (MF) background is the value obtained without reference peptide. By plotting the results of serial dilutions of tested peptides in a semi-logarithmic graph, a straight line could be drawn by linear regression analysis, enabling calculation of the concentration leading to 50% inhibition (IC_{50}).

*Stability of peptide-HLA-A*0201 complexes*

Experiments were performed as described by van der Burg *et al.* (1996). In short, JY cells were treated with emetine (Sigma) to stop protein synthesis and thereby appearance at the cell surface of newly synthesized HLA-A*0201 molecules. Subsequently, HLA-A*0201 complexes were stripped off their peptides by mild acid treatment and reconstituted with the peptide of interest at 200 μg/ml. Peptide-loaded cells were washed with cold IMDM and incubated at 37°C for 0, 2, 4 and 6 hr. The amount of HLA-A*0201-peptide complexes was measured by staining the cells with BB7.2 (supernatant of hybridoma cell culture; ATCC, Rockville, MD), an HLA-A2 conformation-specific antibody and GaM-Fitc (Boehringer, Mannheim, Germany), with subsequent FACScan analysis. The fluorescence index (FI) was calculated for each sample by the formula:

$$\text{FI} = \frac{(\text{MF sample} - \text{MF background})}{\text{MF background}}$$

where MF background is the value without peptide. Samples were tested in duplicate and mean FI was calculated at each time point. The percentage of residual HLA-A2 molecules was calculated by equating the FI at t = 0 to 100% for each peptide, and then using the formula:

$$\%\ \text{remaining}_{t=n} = (\text{FI}_{t=n}/\text{FI}_{t=0}) \times 100\%.$$

As the dissociation of peptides from MHC is a linear process, the stability of the peptides was measured as the time required for 50% of the molecules to decay (DT_{50}) starting from t = 2 hr. By linear regression analysis of the sequential measurements plotted against the % remaining HLA-A2 molecules, the DT_{50} was calculated.

In vivo *immunizations and murine T cell cultures*

Immunogenicity of peptides was determined using HLA-A*0201K^b transgenic mice, kindly provided by Dr. L. Sherman (Scripps Laboratories, San Diego, CA). These transgenic mice express the product of the chimeric *HLA-A*0201*K^b gene, in which only the α3 domain of HLA-A*0201 is replaced by the murine H2-K^b α3 domain, resulting in a molecule that binds HLA-A*0201-binding peptides and interacts with murine $CD8^+$ T cells (Vitiello *et al.*, 1991). The mice were held under clean conventional conditions. Groups of 2–3 mice were injected subcutaneously in the flank with 50 μg peptide mixed with 140 μg HBV core antigen-derived T helper epitope (Millich *et al.*, 1988), emulsified in IFA. Mice were boosted after 14 days with the same mixture. Subsequently, mice were killed 11–14 days after the last immunization, and 3×10^7 spleen cells passed through nylon wool were re-stimulated *in vitro* with 1×10^7 thoroughly washed syngeneic peptide-loaded LPS-elicited lymphoblasts in supplemented IMDM (see Cell culture) in standing T25 tissue-culture flasks. Cultures were incubated for 6 days at 37°C 5% CO_2 in humidified air before their lytic activity was tested.

Cytotoxicity assays (^{51}Cr and Eu release)

CTL activity was measured by standard chromium- or europium-release assays (Visseren *et al.*, 1995). Labeled target cells were loaded with 10 μg/ml peptide for at least 20 min at 37°C. Titrated amounts of effector cells were incubated with constant amounts of target cells for at least 4 hr. Spontaneous and maximal release were measured in multiples of 6. A response was defined as positive when the lysis in a cytotoxicity assay of target cells loaded with the specific peptide was at least 10% higher at 2 E/T ratios than the background lysis of unloaded target cells.

Transfection of COS-7 cells and TNF-release assay

COS-7 cells were transfected with *HLA-A*0201*K^b, *MAGE*-2 and/or tyrosinase cDNA cloned into pcDNAI/amp by DEAE-dextran-chloroquine methods as previously described (Visseren *et al.*, 1995). After 48 hr, the medium was discarded, and the COS-7 cells were used as stimulator cells in a TNF-release assay. In short, 5×10^3 murine bulk culture cells or 2×10^3 human CTL were added to the transfected COS-7 cells. After 24 hr, the supernatant was harvested and its TNF content was determined by its cytolytic effect on mouse fibrosarcoma cell line WEHI 164 clone 13 (WEHI-13), measured by the intensity of vital staining on an ELISA plate reader. The maximum cytolytic effect on WEHI-13 cells was estimated by adding 500 pg/ml human rTNFα (a kind gift of Dr. C. Löwik, Leiden, The Netherlands).

RESULTS

The MAGE-2 protein is expressed in many primary melanomas and their metastases, as well as in other tumors. The possibility of generating CTL that can specifically recognize *MAGE*-2-derived peptides can be of great advantage in the treatment of those tumors. Since HLA-A*0201 is a frequently expressed MHC allele (Imanishi *et al.*, 1992), the presence of HLA-A*0201-restricted CTL epitopes in the MAGE-2 protein was analyzed.

*Selection of peptides and their HLA-A*0201 binding capacity*

Binding of a peptide to a certain MHC class I allele is mainly determined by its length and the fit of the amino acid side chains of the peptide into fixed pockets of the MHC molecule. For HLA-A*0201 binding peptides, both the position and identity of highly important (anchor) and important (dominant) residues in the peptide have been elucidated, resulting in a binding motif (Falk *et al.*, 1991). The amino acid sequence of the *MAGE*-2 gene product was screened for putative HLA-A*0201 binding peptides using a computer scoring program (D'Amaro *et al.*, 1995) utilizing these HLA-A*0201 binding motifs. Peptides fitting the motif (at least 1 anchor and dominant residue or 2 anchors) were synthesized. The HLA-A*0201 binding capacity of all synthesized peptides was tested in a peptide binding competition assay (van der Burg *et al.*, 1996). JY cells (HLA-A*0201 homozygous), expressing empty MHC class I molecules upon acid elution, were incubated with titrated amounts of peptide mixed with a fixed amount of fluorescein-

labeled reference peptide known to bind well to HLA-A*0201. The concentration of the peptides able to inhibit binding of the reference peptide by 50% (IC_{50}) is shown in Table I. Although all 22 peptides tested fit the binding motif, 8 peptides did not inhibit binding of this reference peptide at the highest concentration ($IC_{50} > 100$ μM). Only 7 MAGE-2 peptides, indicated in bold, were capable of inhibiting binding of the reference peptide at low concentrations (less than 20 μM), as is seen for peptides known to bind relatively well to HLA-A*0201 (last 3 rows in Table I) (Millich *et al.*, 1988; Morrison *et al.*, 1992; Wölfel *et al.*, 1994).

Stability of peptide-MHC complexes at physiological temperature

The competition-based peptide binding assay was performed at 4°C, thereby ignoring the influence of higher temperature on stability of the formed complexes. We therefore tested the stability at 37°C of complexes formed with *MAGE*-2-derived peptides of Table I that showed a high or intermediate binding affinity ($IC_{50} < 20$ μM). After acid elution, JY cells were loaded with the peptide of interest. Subsequently, the amount of HLA-A2 molecules at the cell surface was analyzed after various incubation times at 37°C, and the period leading to 50% decay of complexes (DT_{50}) was assessed. Of the 7 peptides that bound well, 3 were able to induce peptide-HLA-A*0201 complexes that showed a DT_{50} of over 6 hr at 37°C: KMVELVHFL (M2 112–120), YLQLVFGIEV (M2 157–166) and KIWEELSMLEV (M2 220–230) (Table II). Interestingly, several peptides that bound with good affinity at 4°C did not form highly stable complexes, as measured at 37°C. These peptides are therefore less likely to form stable complexes *in vivo* and have a lower chance to appear in HLA-A*0201 molecules at the cell surface.

Immunogenicity of MAGE-*2 derived peptides in HLA-A*0201K^b transgenic mice*

To test their immunogenicity *in vivo*, HLA-A*0201K^b transgenic mice were immunized with *MAGE*-2-derived peptides binding with relatively good affinity ($IC_{50} < 20$ μM) to HLA-A*0201. The lytic activity of spleen cells re-stimulated *in vitro* with peptide-loaded syngeneic LPS-elicited lymphoblasts was tested after 1 week of culture. In Figure 1 examples of positive bulk cultures are shown, which indicate that the 3 *MAGE*-2-derived peptides that form the most stable complexes with HLA-A*0201, *i.e.*, M2 112–120 (Fig. 1*a*), M2 157–166 (Fig. 1*b*) and M2 220–230 (Fig. 1*c*), are immunogenic in these transgenic mice. Peptides with relatively good binding affinity but lesser stability at 37°C (M2 44–53, M2 153–161, M2 271–279 and M2 271–281) were not immunogenic. Responses of mice immunized with the known HLA-A*0201-restricted tyrosinase epitope are shown as a control (Fig. 1*d*). As shown in Table II, we only found responses to the 3 MAGE-2 peptides that form stable complexes with HLA-A*0201, and the 2 control peptides.

Recognition of cells expressing MAGE *-2 and HLA-A*0201K^b*

The peptide-specific bulk cultures mentioned above were again re-stimulated with peptide-loaded spleen cells, and recognition of COS-7 cells transfected with the proper HLA gene and the gene coding the antigen was tested. Recognition of COS-7 cells transfected with HLA-A*0201K^b and tyrosinase by bulk cultures derived from mice immunized with tyrosinase 369–377 peptide is shown as a control (Table III). The bulk cultures from mice immunized with M2 220–230 did not yield enough cells to perform this experiment. As shown in Table III, the bulk cultures derived from mice immunized with M2 112–120 (mice 1–3) and M2 157–166 peptides (mice 5–6) showed recognition of COS-7 cells transfected with *HLA-A*0201K^b* and *MAGE*-2, indicating that these 2 peptides are processed and presented in HLA-A*0201.

TABLE I – BINDING AFFINITIES OF SELECTED *MAGE*-2 DERIVED PEPTIDES TO HLA-A*0201

Position	AA sequence	IC_{50}[1]
M2 15–23	GLEARGEAL	>100
M2 24–32	GLVGAQAPA	30
M2 43–53	STLVEVTLGEV	>100
M2 44–53	TLVEVTLGEV	**17**
M2 45–53	LVEVTLGEV	>100
M2 112–120	KMVELVHFL	**7**
M2 149–158	VIFSKASEYL	>100
M2 153–161	KASEYLQLV	**14**
M2 157–166	YLQLVFGIEV	**7**
M2 159–167	QLVFGIEVV	47
M2 159–169	QLVFGIEVVEV	26
M2 171–179	PISHLYILV	80
M2 181–189	CLGLSYDGL	42
M2 194–204	QVMPKTGLLII	82
M2 198–206	KTGLIIVL	27
M2 203–211	IIVLAIIAI	>100
M2 220–230	KIWEELSMLEV	**10**
M2 271–279	FLWGPRALI	**6**
M2 271–281	FLWGPRALIET	**9**
M2 277–286	ALIETSYVKV	>100
M2 278–286	LIETSYVKV	>100
M2 278–287	LIETSYVKVL	>100
HBV 18–27[2]	FLPSDDFPSV	1
INF A M 58–66	GIGLFVFTL	3
TYR 369–377	YMNGTMSQV	9

[1]Concentration of indicated peptide in μM that inhibits binding of the reference peptide to 50%. Bold, less than 20 μM.–[2]Control peptides: HBV core positions 18–27 (Millich *et al.*, 1988); influenza A matrix protein positions 58–66 (Morrison *et al.*, 1992); tyrosinase protein positions 369–377 (Wölfel *et al.*, 1994).

TABLE II – STABILITY AND IMMUNOGENICITY OF PEPTIDES BINDING TO HLA-A*0201 WITH HIGH AND INTERMEDIATE AFFINITY

Position	AA sequence	IC_{50}[1]	DT_{50}[2]	Immunogenicity[3]
M2 44–53	TLVEVTLGEV	17	2	0/5
M2 112–120	KMVELVHFL	7	>4	**3/5**
M2 153–161	KASEYLQLV	14	1.5	0/7
M2 157–166	YLQLVFGIEV	7	>4	**4/7**
M2 220–230	KIWEELSMLEV	10	>4	**2/7**
M2 271–279	FLWGPRALI	6	1	0/7
M2 271–281	FLWGPRALIET	9	3	0/5
INF A M 58–66	GIGLFVFTL	3	>4	**6/7**
TYR 369–377	YMNGTMSQV	9	>4	**2/3**

[1]See Table I.–[2]Hours required for 50% of the MHC-peptide complexes to decay, starting from t = 2 hr at 37°C. Positive immunogenicity indicated in bold.–[3]Number of positive bulk cultures from the total amount of bulk cultures tested in several immunization experiments. Positive bulk cultures showed at least 10% difference in specific lysis of peptide-loaded target cells in the 2 highest effector to target ratios.

DISCUSSION

Although therapy of cancer patients has markedly improved over the last decades, treatment of several types of disseminated cancer has met with limited success. One problem with current treatments of disseminated cancer in general is their low specificity for tumor cells. Most tumor cells differ from normal tissue only by (over)expression of oncogene products or slight mutations in these and renewed expression of embryonal proteins. Such subtle differences can, however, be seen by CTLs, when they lead to presentation of new (tumor-specific) MHC class I binding peptides, or abnormally high levels of MHC class I binding peptides.

MAGE-2 is likely to be such an aberrantly expressed embryonal protein (De Backer *et al.*, 1995), which is present in 70% of metastatic melanomas, and in an appreciable percentage of other tumors such as laryngeal tumors, lung tumors and sarcomas. CTLs recognizing MAGE-2 thus encompass considerable therapeutic potential. In this study we identified 7 *MAGE*-2-derived peptides that bind to HLA-A*0201 with relatively good affinity. Three of these peptides form highly stable complexes with HLA-A*0201 at 37°C. We immunized HLA-A*0201K^b transgenic mice with all *MAGE*-2-derived peptides binding to HLA-A*0201 with good

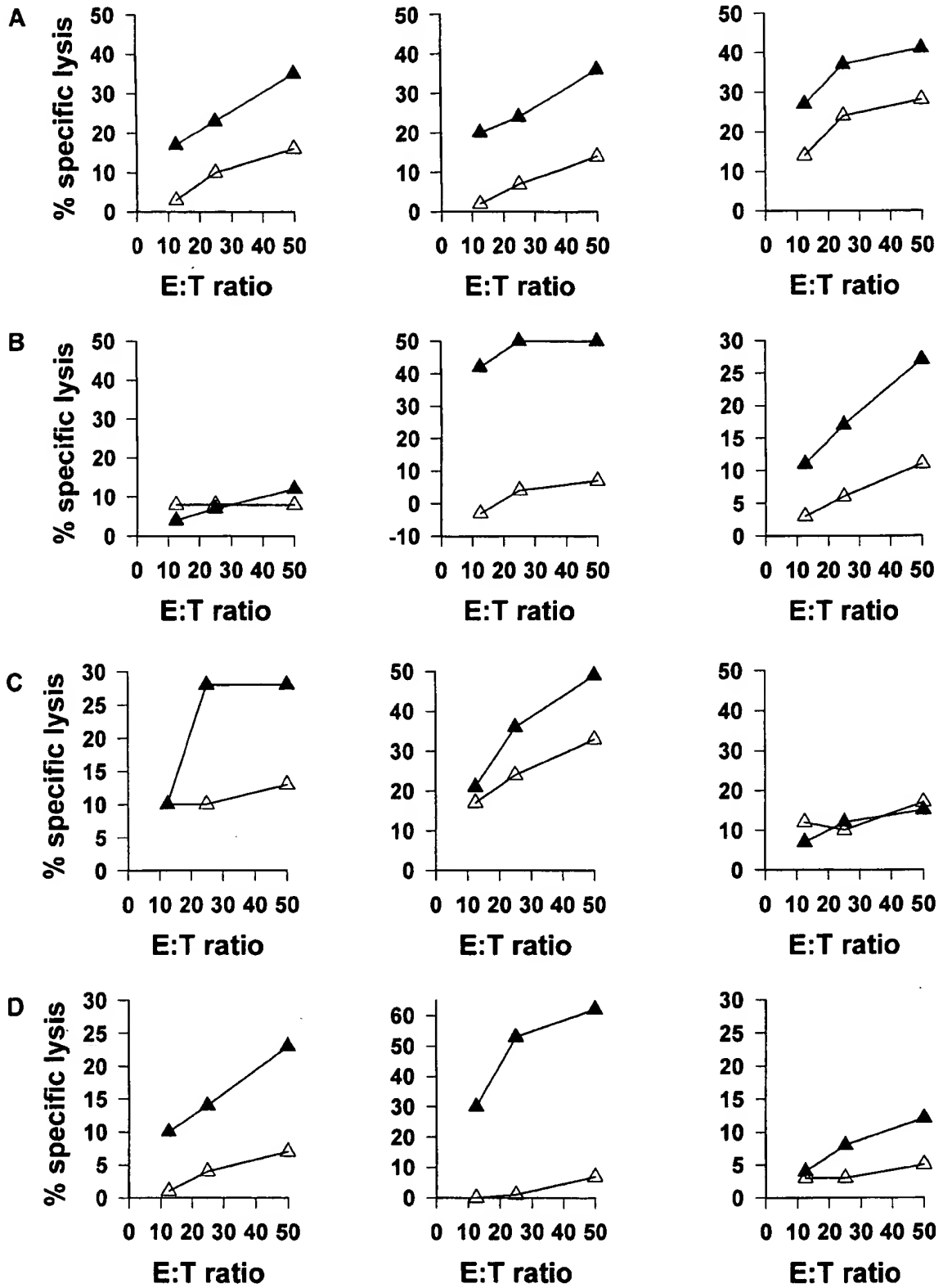


FIGURE 1 – Immunogenicity of *MAGE*-2-derived peptides in HLA-A*0201K^b transgenic mice. Bulk CTL obtained from immunized mice were tested for lytic activity using chromium-labeled Jurkat A*0201K^b target cells (open triangles) or Jurkat A*0201K^b cells pre-incubated with 10 μM relevant peptide (closed triangles). Examples of 3 observed responses are shown. *(a)* Mice immunized with M2 112-120 peptide. *(b)* Mice immunized with M2 157-166. *(c)* Mice immunized with M2 220-230. *(d)* Mice immunized with tyrosinase 369-377 peptide.

TABLE III – RECOGNITION OF ENDOGENOUSLY PROCESSED AND PRESENTED MAGE-2 AND TYROSINASE BY MOUSE CTL BULK CULTURES

Responders[1]	Vaccine[2]	Stimulator cells		
		COS	COS + A*0201K^b	COS[3] + A*0201K^b + antigen
Mouse 1	M2 112–120	<1[4]	18	**90**
Mouse 2	M2 112–120	<1	5	**63**
Mouse 3	M2 112–120	<1	17	**42**
Mouse 4	M2 157–166	<1	11	16
Mouse 5	M2 157–166	<1	2	**26**
Mouse 6	M2 157–166	<1	<1	**11**
Mouse 7	tyr 369–377	<1	<1	**28**
Mouse 8	tyr 369–377	<1	<1	**53**
Mouse 9	tyr 369–377	<1	<1	<1
LB705	M3 271–279	<1	<1	**100**

[1]Responding cells are bulk cultures from immunized HLA-A*0201K^b transgenic mice. LB705 is a human T cell clone *in vitro* induced with a *MAGE*-3-derived peptide that cross-reacts to MAGE-2.–[2]HLA-A*0201K^b transgenic mice were immunized with the indicated peptide. Spleen cells were restimulated twice *in vitro* with peptide-loaded spleen cells at weekly intervals. One week after the last restimulation bulk cultures were tested.–[3]Cells from mice 1–6 and LB705 were tested on *MAGE*-2 cDNA and *HLA-A*0201K^b* -transfected COS-7 cells; bulk cultures from mice 7–9 were tested on COS-7 cells transfected with tyrosinase cDNA and *HLA-A*0201K^b*. –[4]Percentage of WEHI cell death. Background WEHI cell death was subtracted. Positive recognition indicated in bold.

affinity. Only the peptides that were able to induce highly stable HLA-A*0201-peptide complexes were immunogenic in HLA-A*0201K^b transgenic mice, confirming our results on the importance of stability of peptide-MHC complexes for immunogenicity (van der Burg *et al.*, 1996).

Furthermore, we studied the natural processing and presentation of the immunogenic peptides in cells transfected with both *MAGE*-2 and *HLA-A*0201K^b*. Whether or not M2 220–230 (KIWEELSMLEV) is processed and presented remains to be established, since the bulk cultures specific for this peptide were lost before they could be tested. However, the peptides M2 112–120 (KMVELVHFL) and M2 157–166 (YLQLVFGIEV) are processed and presented in cells expressing HLA-A*0201 and MAGE-2. Bulk cultures specific for those peptides recognized COS-7 cells transfected with both *MAGE*-2 and *HLA-A*0201K^b*. No recognition of HLA-A*0201$^+$ melanoma cell lines could be detected (data not shown). Clearly, the human HLA-A*0201$^+$ melanoma cell lines do not express the murine MHC class I α3-domain, causing less optimal conditions for recognition by murine CD8$^+$ CTLs. The lack of recognition of HLA-A*0201-expressing cells by CTLs derived from peptide-vaccinated HLA-A*0201K^b transgenic mice has been reported (Theobald *et al.*, 1995). Therefore, the absence of recognition of human HLA-A*0201-positive MAGE-2-expressing tumor cells by murine CTLs does not exclude recognition of the naturally processed and presented MAGE-2 peptides in the proper HLA molecule.

MAGE-2 is not expressed in normal adult tissues, with the exception of testis tissue (De Plaen *et al.*, 1994). The presence of CTLs specific for MAGE-1 and MAGE-3, which display similar expression patterns compared with MAGE-2, has been demonstrated previously. Therefore, it is highly unlikely that MAGE-2-specific CTLs are absent due to clonal deletion. Furthermore, the existence of autoreactive tissue-specific CTL in healthy donors has been demonstrated (Visseren *et al.*, 1995), implicating their potential use in cancer immunotherapy (Pardoll, 1994).

Although MAGE-2 is present in the majority of melanomas, so far no HLA-A*0201-restricted MAGE-2-specific CTLs have been found in melanoma patients. This indicates that other antigens, such as gp100/pMel17 (Bakker *et al.*, 1994), tyrosinase (Wölfel *et al.*, 1994) and Melan A/MART-1 (Coulie *et al.*, 1994), are more immunogenic than MAGE-2 in HLA-A*0201$^+$ patients. Thus, inducing MAGE-2-specific CTLs in these patients would add a new specificity of a subdominant nature to the anti-tumor CTLs already present. The usefulness of immunosubdominant epitope-based peptide vaccines for prevention and treatment of tumors has been demonstrated (Feltkamp *et al.*, 1993). A broad anti-tumor CTL repertoire would significantly reduce the opportunities for tumor cells to escape from CTL attack by altering CTL epitopes. Vaccination of cancer patients with epitope-based vaccines containing several CTL epitopes matched with their HLA type could therefore enhance the chances of therapeutic success.

ACKNOWLEDGEMENTS

We thank Mr. G. Schijff for handling and keeping the transgenic mice, Dr. J.W. Drijfhout and Miss W. Benckhuijsen for synthesizing the peptides and Drs. T. Ottenhof and R. Offringa for critically reading the manuscript.

This study was supported by a grant from the Macropa Stichting. Both MV and EvdV are supported by the Dutch Cancer Society (grant NKB 91-15).

REFERENCES

BAKKER, A.B.H., SCHREURS, M.W.J., DE BOER, A.J., KAWAKAMI, Y., ROSENBERG, S.A., ADEMA, G.J. and FIGDOR, C.G., Melanocyte lineage-specific antigen *gp100* is recognized by melanoma-derived tumor-infiltrating lymphocytes. *J. exp. Med.*, **179**, 1005–1009 (1994).

COULIE, P.G., BRICHARD, V., VAN PEL, A., WÖLFEL, T., SCHNEIDER, J., TRAVERSARI, C., MATEI, S., DE PLAEN, E., LURQUIN, C., SZIKORA, J.-P., RENAULD, J.-C. and BOON, T., A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. *J. exp. Med.*, **180**, 35–42 (1994).

D'AMARO, J., HOUBIERS, J.G.A., DRIJFHOUT, J.-W., BRANDT, R.M.P., SCHIPPER, R., BOUWES BAVINCK, J.N., MELIEF, C.J.M. and KAST, W.M., A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide binding motifs. *Hum. Immunol.*, **43**, 13–18 (1995).

DE BACKER, O., VERHEYDEN, A.-M., MARTIN, B., GODELAINE, D., DE PLAEN, E., BRASSEUR, R., AVNER, P. and BOON, T., Structure, chromosomal location and expression pattern of three mouse genes homologous to the human *MAGE* genes. *Genomics*, **28**, 74–83 (1995).

DE PLAEN, E. and 14 OTHERS, Structure, chromosomal localization and expression of twelve genes of the *MAGE* family. *Immunogenetics*, **40**, 360–369 (1994).

DE SMET, C., LURQUIN, C., VAN DER BRUGGEN, P., DE PLAEN, E., BRASSEUR, F. and BOON, T., Sequence and expression pattern of expression of human gene *MAGE*-2. *Immunogenetics*, **39**, 121–129 (1994).

FALK, K., RÖTSCHKE, O., STEVANOVIC, S., JUNG, G. and RAMMENSEE, H.G., Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature (Lond.)*, **351**, 290–296 (1991).

FELTKAMP, M.C.W., SMITS, H.L., VIERBOOM, M.P.M., MINNAAR, R.P., DE JONGH, B.M., DRIJFHOUT, J.W., TER SCHEGGET, J., MELIEF, C.J.M. and KAST, W.M., Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. *Europ. J. Immunol.*, **23**, 2242–2249 (1993).

IMANISHI, T., AKAZA, T. and KIMURA, A., Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. *In:* Tsuji K., M. Aizowa, and T. Sasazuki (eds.), *HLA 1991, Proceedings of the Eleventh International Histocompatibility Workshop and Conference*, pp. 1066–1077, Oxford University Press (1992).

MELIEF, C.J.M. and KAST, W.M., Lessons from T cell responses to virus-induced tumors for cancer eradication in general. *Cancer Surv.*, **13**, 81–99 (1992).

MILLICH, D.R., HUGHES, J.L., MCLACHLAN, A., THORNTON, G.B. and MORIARTY, A., Hepatitis B synthetic immunogen comprised of nucleocapsid T-cell sites and an envelope B-cell epitope. *Proc. nat. Acad. Sci. (Wash.)*, **85**, 1610–1615 (1988).

MIZOGUCHI, H., O'SHEA, J.J., LONGO, D.L., LOEFFLER, C.M., MCVICAR, D.W., and OCHOA, A.C., Alteration in signal transduction molecules in T lymphocytes from tumor bearing mice. *Science,* **158,** 1795–1798 (1992).

MORRISON, J., ELVIN, J., LATRON, F., GOTCH, F., MOOTS, R., STROMINGER, J.L. and MCMICHAEL, A., Identification of the nonamer peptide from influenza A matrix protein and the role of pockets of HLA-A2 in its recognition by cytotoxic T lymphocytes. *Europ. J. Immunol.,* **22,** 903–907 (1992).

PARDOLL, D.M., A new look for the 1990s. *Nature (Lond.),* **369,** 357–358 (1994).

RUITER, D.J., BERGMAN, W., WELVAART, K., SCHEFFER, E., VAN VLOTEN, W.A., RUSSO, C. and FERRONE, S., Immunohistochemical analysis of malignant melanomas and nevocellular nevi with monoclonal antibodies to distinct monomorphic determinants of HLA antigens. *Cancer Res.,* **44,** 3930–3935 (1984).

THEOBALD, M., BIGGS, J., DITTMER, D., LEVINE, A.J. and SHERMAN, L.A., Targeting p53 as a general tumor antigen. *Proc. nat. Acad. Sci. (Wash.),* **92,** 11993–11997 (1995).

TORRE-AMIONE, G., BEAUCHAMP, R.D., KOEPPEN, H., PARK, B.H., SCHREIBER, H., MOSES, H.L. and ROWLEY, D.A., A highly immunogenic tumor transfected with a murine transforming growth factor type beta 1 cDNA escapes immune surveillance. *Proc. nat. Acad. Sci. (Wash.),* **87,** 1486–1490 (1990).

TOWNSEND, A.R.M., ROTHBARD, J., GOTCH, F.M., BAHADUR, G., WRAITH, D. and MCMICHAEL, A.J., The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. *Cell,* **44,** 959–968 (1986).

TRAVERSARI, C., VAN DER BRUGGEN, P., VAN DEN EYNDE, B., HAINAUT, P., LEMOINE, C., OHTAK, N., OLD, L. and BOON, T., Transfection and expression of a gene coding for a human melanoma antigen recognized by autologous CTL. *Immunogenetics,* **35,** 145–152 (1992).

VAN DER BRUGGEN, P., BASTIN, J., GAJEWSKI, T., COULIE, P.G., BOËL, P., DE SMET, C., TRAVERSARI, C., TOWNSEND, A. and BOON, T., A peptide encoded by human gene *MAGE*-3 and presented by HLA-A2 induces cytolytic T lymphocytes that recognize tumor cells expressing *MAGE*-3. *Europ. J. Immunol.,* **24,** 3038–3043 (1994).

VAN DER BURG, S.H., VISSEREN, M.J.W., BRANDT, R.M.P., KAST, W.M. and MELIEF, C.J.M., Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J. Immunol.,* **156,** 3308–3314 (1996).

VISSEREN, M.J.W., KOOT, M., VAN DER VOORT, E.I.H., GRAVESTEIN, L.A., SCHOENMAKERS, H.J., KAST, W.M., ZIJLSTRA, M. and MELIEF, C.J.M., Production of interleukin-2 by EL4 tumor cells induces natural killer cell- and T-cell-mediated immunity. *J. Immunother.,* **15,** 119–128 (1994).

VISSEREN, M.J.W., VAN ELSAS, A., VAN DER VOORT, E.I.H., RESSING, M.E., KAST, W.M., SCHRIER, P.I. and MELIEF, C.J.M., CTL specific for the tyrosinase autoantigen can be induced from healthy donor blood to lyse melanoma cells. *J. Immunol.,* **154,** 3991–3998 (1995).

VITIELLO, A., MARCHESINI, D., FURZE, J., SHERMAN, L.A. and CHESNUT, R.W., Analysis of the HLA-restricted Influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex. *J. exp. Med.,* **173,** 1007–1015 (1991).

WÖLFEL, T., VAN PEL, A., BRICHARD, V., SCHNEIDER, J., SELIGER, B., MEYER ZUM BÜSCHENFELDE, K.H. and BOON, T., Two tyrosinase nonapeptides recognized on HLA-A2 melanomas by autologous cytolytic T lymphocytes. *Europ. J. Immunol.,* **24,** 759–764 (1994).